

# Acoustic and Perceptual Analyses of Politeness in Japanese Speech

Etsuko Ofuka  
✓

Submitted in accordance with the requirements for the degree of  
Doctor of Philosophy

The University of Leeds  
Department of Psychology

March 1996

The candidate confirms that the work submitted is her own and that appropriate  
credit has been given where reference has been made to the work of others.

## ABSTRACT

In order to examine potential acoustic cues for politeness in Japanese speech,  $f_0$  and temporal aspects of polite and casual utterances of two question sentences spoken by six male native speakers were acoustically analysed. The analysis showed that  $f_0$  movement of the final part of utterances and speech rate of utterance were consistently differently used in these different speaking styles (i.e., 'polite' and 'casual') across all the speakers. Perceptual experiments with listeners using a rating scale method confirmed that these acoustic variables, which were manipulated using digital resynthesis, had an impact on politeness perception. It was showed that the duration and  $f_0$  direction of the final vowel of utterances were so influential that the overall impression of utterance politeness was changed. An experiment which used speech rate variations of a polite utterance showed the important role of this variable in perceived politeness. Politeness ratings showed an inverted-U shape as a function of speech rate, but differed according to particular speakers. The speech rate of listeners was found to affect their utterance rate preference; listeners clearly preferred rates close to their own, i.e., rates they perceived as 'natural' or comfortable. A final experiment, using speech rate variations of a polite utterance as stimuli and a two alternative forced-choice procedure, showed a very high correlation between perceived politeness scores and naturalness scores. This suggests the importance of listener characteristics in politeness research.

## CONTENTS

<b>LIST OF FIGURES</b>	<b>vii</b>
<b>LIST OF TABLES</b>	<b>x</b>
<b>ACKNOWLEDGEMENTS</b>	<b>xiii</b>

<b>CHAPTER</b>	<b>PAGE</b>
<b>1. INTRODUCTION</b>	<b>1</b>
1.1 Why is 'polite' prosody so important?	1
1.2 Organisation of the thesis	4
<b>2. JAPANESE POLITENESS</b>	<b>5</b>
2.1 What is politeness?	5
2.1.1 The concept of politeness	5
2.1.2 Politeness as a strategy	6
2.1.3 Western politeness vs Japanese politeness	7
2.2 'Keigo' in use in the future	12
2.3 Empirical studies of 'keigo'	14
2.4 Summary	16
<b>3. RESEARCH ON PARALINGUISTIC CUES TO SPEAKER VARIABLES</b>	<b>17</b>
3.1 Introduction	17
3.2 What do 'prosodic' and 'paralinguistic' mean?	17
3.3 Potential paralinguistic cues to speaker variables	18
3.3.1 Introduction	18
3.3.2 Potential paralinguistic cues to Japanese politeness	18
3.3.3 Acoustic cues to speaker variables	21
3.3.3.1 Acoustic variables related to pitch	22
3.3.3.2 Temporal variables	25
3.3.3.3 Acoustic variables related to loudness	27
3.3.3.4 Acoustic variables related to voice quality and	

articulation.....28

3.4 Methodological issues regarding speech data collection.....30

3.4.1 Elicitation methods.....30

3.4.2 Selection of test passages.....32

3.4.3 Speaker variability.....33

3.5 Methodological issues regarding perceptual experiments.....33

3.5.1 Acoustic cue manipulation.....34

3.5.1.1 Synthesis/resynthesis techniques.....34

3.5.1.1.1 Articulatory synthesis technique.....34

3.5.1.1.2 Formant synthesis technique.....35

3.5.1.1.3 Linear Prediction (LP) analysis-synthesis  
technique.....36

3.5.1.1.4 Pitch Synchronous Overlap and Add (PSOLA)  
technique.....37

3.5.1.2 Considerations on selection of techniques.....38

3.5.1.3 Ecological validity problems.....39

3.5.2 Rating tasks.....43

3.5.2.1 Rating methods and stimulus presentation.....43

3.5.2.2 Stimuli.....45

3.5.2.3 Context effects.....46

3.5.2.4 Listener-judges.....47

3.5.2.5 Statistical considerations.....48

3.6 Summary.....51

3.7 Approach adopted in the present study.....52

**4. SPEECH DATA COLLECTION.....53**

4.1 Politeness elicitation.....53

4.2 Utterance evaluation.....57

4.2.1 Method.....57

4.2.2 Results and discussion.....59



<b>5. ACOUSTIC ANALYSIS.....</b>	<b>62</b>
5.1 Acoustic features chosen for measurement.....	62
5.2 The outcome of acoustic analyses.....	64
5.2.1 F0 level.....	71
5.2.2 F0 variability, range and rate of change.....	73
5.2.3 Articulation rate.....	74
5.2.4 Total utterance length and pause.....	74
5.2.5 Final f0 movement.....	76
5.2.6 Differences in voice quality and articulation.....	77
5.3 Summary.....	79
<b>6. PERCEPTUAL EXPERIMENTS.....</b>	<b>84</b>
6.1 Experiments on the role of the final part of utterances....	85
6.1.1 Experiment 1-A: The effects of final f0 movement.....	85
6.1.1.1 Method.....	85
6.1.1.2 Results and discussion.....	88
6.1.2 Experiment 2: The effects of speaking style of the final part.....	99
6.1.2.1 Method.....	99
6.1.2.2 Results and discussion.....	104
6.1.3 Potential effects of f0 manipulation on voice quality.....	112
6.1.3.1 Vowel production.....	112
6.1.3.2 Spectral analysis and auditory evaluation.....	114
6.1.4 Summary of the experiments on the final part.....	117
6.2 Experiment 1-B: The role of speech rate.....	118
6.2.1 Pre-test: Listening test for assessing the normality of speech rates.....	119
6.2.1.1 Procedure.....	119
6.2.1.2 Results.....	120

6.2.2 Experiment 1-B.....	122
6.2.2.1 Method.....	122
6.2.2.2 Results and discussion.....	123
6.3 Experiment 3: The role of perceived naturalness.....	131
6.3.1 Method.....	131
6.3.2 Results and discussion.....	133
6.4 Summary.....	142
<b>7. GENERAL CONCLUSIONS.....</b>	<b>144</b>
7.1 Summary of the findings.....	144
7.2 Future work.....	147
<b>REFERENCES.....</b>	<b>149</b>
 <b>APPENDIX 1.....</b>	 <b>162</b>
<b>APPENDIX 2.....</b>	<b>175</b>
<b>APPENDIX A.....</b>	<b>198</b>
<b>APPENDIX B.....</b>	<b>206</b>
<b>APPENDIX C.....</b>	<b>211</b>
<b>APPENDIX D.....</b>	<b>224</b>
<b>APPENDIX E.....</b>	<b>232</b>
<b>APPENDIX F.....</b>	<b>238</b>
<b>APPENDIX G.....</b>	<b>242</b>
<b>APPENDIX H.....</b>	<b>253</b>
<b>APPENDIX I.....</b>	<b>249</b>
<b>APPENDIX J.....</b>	<b>254</b>

## LIST OF FIGURES

FIG. 4.1. Components of situation.....	55
FIG. 5.1. Wide bandwidth spectrograms of 'moshimoshi' spoken by KS.....	80
FIG. 5.2. Narrow bandwidth spectrograms of 'moshimoshi' spoken by KS.....	81
FIG. 5.3. Wide bandwidth spectrograms of 'moshimoshi' spoken by TK.....	82
FIG. 5.4. Narrow bandwidth spectrograms of 'moshimoshi' spoken by TK.....	83
FIG. 6.1. Rating scale of politeness used in Experiment 1.....	87
FIG. 6.2. Comparisons between the mean values of politeness scores for the short final duration (SHORT) and long final duration (LONG) versions of the utterances originally spoken by KS and TK in Experiment 1-A.....	92
FIG. 6.3. Comparisons between the mean values of politeness scores for the final rise (RISE) and final fall (FALL) versions of the utterances originally spoken by KS and TK in Experiment 1-A.....	94
FIG. 6.4. Mean values of politeness scores rated by 20 subjects for the polite/casual source utterances with the 'matched' and 'conflict' final prosody in Experiment 1-A.....	96
FIG. 6.5. Style preferences of 20 subjects for polite (i.e., the polite style with the 'preferred' final prosody) and casual (i.e., the casual style with the 'less preferred' final prosody) utterances spoken by two speakers, according to the accent of the subjects in Experiment 1-A.....	98

FIG. 6.6. Schematic figures of the actual final f0 movements....	101
FIG. 6.7. Rating scales for anger, kindness, politeness and naturalness used in Experiment 2.....	103
FIG. 6.8. Mean ratings for politeness by speaking style of the first part (First Style), speaking style of the final mora (Final Style) and prosody of the final mora (Final Prosody) in Experiment 2.....	109
FIG. 6.9. Mean politeness and naturalness scores for each condition in Experiment 2.....	111
FIG. 6.10. Schematic figures of f0 movements of two vowels with a level tone and a rising tone spoken by one male speaker.....	113
FIG. 6.11. Estimated formant structures for two vowels: 'ae' (a) and 'ah' (b).....	115
FIG. 6.12. Scale used in the pre-test.....	119
FIG. 6.13. Mean values of politeness scores across 20 subjects for five different rate versions of the sentence originally spoken by KS and TK in Experiment 1-B.....	126
FIG. 6.14. Rate preferences of 12 male subjects in Experiment 1-B as a function of the rate of the utterances.....	129
FIG. 6.15. Rate preferences of 7 female subjects in Experiment 1-B as a function of the rate of the utterances.....	130
FIG. 6.16. A male subject (M1)'s politeness scores and naturalness scores for the utterances by TK in Experiment 3.....	134
FIG. 6.17. A male subject (M2)'s politeness scores and naturalness	



scores for the utterances by TK in Experiment 3.....135

FIG. 6.18. A female subject (F1)'s politeness scores and naturalness  
scores for the utterances by TK in Experiment 3.....136

FIG. 6.19. A female subject (F2)'s politeness scores and naturalness  
scores for the utterances by TK in Experiment 3.....137

FIG. 6.20. Mean politeness and naturalness scores for the four  
subjects for the five different rate versions in  
Experiment 3.....139

FIG. 6.21. Mean politeness and naturalness scores for the four  
subjects in Experiment 3.....140

## LIST OF TABLES

TABLE 4.1. Age and hometown of the speakers.....	56
TABLE 4.2. Kendall's coefficient of concordance (W) between five listener-judges' scores for six different speaker conditions.....	59
TABLE 4.3. Mean scores across five listener-judges' scores in the utterance evaluation test.....	61
TABLE 5.1. F0 variables in polite (P) and casual (C) utterances: mean values and SDs across six male speakers.....	66
TABLE 5.2. Temporal variables in polite (P) and casual (C) utterances: mean values and SDs across six male speakers.....	67
TABLE 5.3. Final morae: duration and f0 rate of change in Phrase 1 and Phrase 2 of polite (P) and casual (C) utterances..	68
TABLE 5.4. Comparisons between f0 and temporal aspects of polite versions (P) and those of casual versions (C) of two sentences spoken by six male speakers.....	69
TABLE 5.5. Comparisons between the characteristics of the final mora of the first and the second phrase of the sentence of polite versions (P) and those of casual versions (C) of two sentences spoken by six male speakers.....	70
TABLE 5.6. Pause durations (in ms) in three different speaking styles in two sentences (the 'luggage' sentence (a) and the 'hello' sentence (b)) spoken by six male speakers.....	75
TABLE 6.1. ANOVA results of Experiment 1-A: significant effects	

at the level of 0.05 or better.....	90
TABLE 6.2. Mean politeness ratings with standard deviations (SD) in Experiment 1-A.....	91
TABLE 6.3. Duration and f0 characteristics of the final mora ('ka') in each condition.....	102
TABLE 6.4. Mean values and SDs over 19 subjects' five repetition scores for politeness, anger, kindness and reaction time.....	105
TABLE 6.5. ANOVA results of Experiment 2: significant effects at the level of 0.05 or better.....	107
TABLE 6.6. Perceived tempo index (PTX) and speech rate (SR) for the nine speech rate levels (S30 ~ F30) of the polite versions of the 'luggage' sentence spoken by three male speakers (KS, TK and SF).....	121
TABLE 6.7. ANOVA results of Experiment 1-B: significant effects at the level of 0.05 or better.....	124
TABLE 6.8. Mean politeness ratings with standard deviations (SD) in Experiment 1-B.....	125
TABLE 6.9. Speech rates of subjects.....	127
TABLE 6.10. Inter-trial agreement among each subject's five trial blocks for the five different rate versions of the sentence originally spoken by the two speakers (KS and TK) for politeness and naturalness in Experiment 3, using Kendall's coefficient of concordance (W) with a level of significance better than 0.01.....	138
TABLE 6.11. Correlations between politeness scores and	

naturalness scores of the five different rate versions of the sentence originally spoken by the two speakers (KS and TK) by the Pearson product-moment correlation coefficient ( $r$ ) in Experiment 3.....	141
---	-----



## ACKNOWLEDGEMENTS

Since I started this work, I have received help and encouragement from so many people in various ways, and it is almost impossible to mention them all here. But my special thanks must go to my supervisors, Professor Peter Roach, Dr. Denis McKeown and Dr. Mitch Waterman, for all the help, guidance and encouragement they gave me. I am also very grateful to Dr. Celia Scully for her wonderful lectures on experimental phonetics, which fascinated me and eventually made me start working in the area of speech science. Without them, this work would have never been initiated and completed.

I must also thank Akiko Kawasaki, Professor Sachiko Ide, Professor Tsunao Ogino and Dr. Hong MinPyo for useful discussions and guidance about politeness theories. I am especially grateful to Akiko Kawasaki for her invaluable encouragement and help including sending me relevant articles and books published in Japan and giving me opportunities to do perceptual experiments with her students. I also thank Dr. Masako Yuasa for her help with preparation of scenarios and recording procedures.

Dr. Nick Campbell and Dr. Yoshinori Sagisaka gave me opportunities to use their speech materials and facilities. I would like to thank them and all members of Department 2, ITL/ATR for their valuable suggestions and discussions. I especially thank Dr. Hélène Valbret for her technical support.

I owe thanks to all my friends and colleagues for their useful discussions and technical support. Peter Greasley and Jane Setter patiently read my drafts carefully and helped me in improving my English writing. Dave Horton, Simon Arnfield, Karen Stromberg and John Blinkhorn gave me technical support. Rob Donovan and Shigeto Furukawa answered all my questions about speech synthesis

techniques and signal detection theories respectively.

Finally, I sincerely thank all the speakers and subjects who kindly participated in the recording and rating sessions. Since most of the work was done in England, one of the most difficult parts was to find native speakers of Japanese who were willing to be our subjects. Many people have helped me. Among them, I especially thank Megumi Fujioka, Noriko Kawakami, Ken Sotowa and Machiko Sato.

# CHAPTER 1

## INTRODUCTION

### 1.1. Why is 'polite' prosody so important?

In any language community, the speaker's ability to show an appropriate level of politeness is very important for smooth social interaction. It is especially so for the Japanese speech community, because Japanese society still attaches much importance to the hierarchy of social relationship: Japanese society has been described as a 'vertical society' (Nakane, 1967, 1970) depending on various factors such as age, sex, and social status, and appropriate use of politeness actually acknowledges and maintains the social hierarchy (Matsumoto, 1988). In the Japanese language, the level of politeness is encoded in a special linguistic system called *keigo* ('honorific system'), having special expressions or words for displaying respect and modesty, as well as non-verbal forms such as body and facial expressions, tone of voice, and appearance.

Since the linguistic forms, attitudes, and appearance have been regarded as very important in conveying politeness, there are various kinds of textbooks and formal teaching at school and also in the work place in Japan. However, little attention has been given to teaching how to make utterances sound polite, although tone of voice is known to be important. This neglect may reflect the fact that the Japanese culture has valued silence much more than speech, evoked by the saying "silence is golden". However, the rapid increase in the number of households and persons owning a telephone and the importance of telephones especially in business have been changing people's attitudes towards speech: the importance of speech, and therefore the importance of learning how to speak is beginning to be recognised. Nippon Housou Kyoukai (the Japanese state broadcasting institute) has recently



started to broadcast a series of educational programs on the radio with the focus on various aspects of spoken language including the right usage of the honorific system, good pronunciation and speaking styles, and good manners for telephone conversations (NHK, 1995).

The present study focuses on speech in relation to politeness perception. In fact, the importance of how to speak in terms of politeness has been recognised by native speakers together with the importance of the linguistic forms. Ogino and Hong (1992) conducted a questionnaire survey on what cues Japanese people would use to evaluate the level of politeness of the speaker, with more than 200 Tokyo residents between 23 and 74 years of age taking part. This survey showed that a Japanese person would mostly rely on the appropriateness of the speaker's use of the honorific system, followed by facial expressions, tone of voice, gaze, gesture, and clothes or shoes.

Hong (1992) also mentioned the importance of research on the way of speaking in terms of conveying politeness from the point of view of teaching Japanese as a foreign language. He presented over 100 native speakers of Japanese with polite utterances spoken by six native speakers and the same sentences spoken by six Korean learners of Japanese, and asked them to judge whether the utterances sounded polite or not to them. The results showed that the polite utterances spoken by Korean learners were perceived as polite by no more than half of the native listeners, while the native utterances were appropriately identified by more than 80% of listeners. He concluded that this was probably due to the incorrect prosody imposed on the utterances by the learners. In fact, this is a serious concern for learners of foreign languages. I became acquainted with a number of learners of Japanese who were very keen to learn how to express politeness and familiarity properly in Japanese. Many of them expressed irritation with their inability to express familiarity in a foreign language. This feeling is very familiar to any



learners of foreign languages, including myself as a learner of English. Learning languages is not just learning how to construct grammatically correct sentences. We have to know how to speak those sentences appropriately in a given situation.

Given the likely importance of prosody as a politeness cue, listeners would be expected to be highly sensitive to the acoustic variables underlying prosody. If we wish to study the effects of manipulated acoustic variables on politeness perception, we need to ensure that listeners' sensitivity extends to such manipulated variables.

In a pilot experiment (reported in detail in Appendix 1) it is demonstrated that synthetic speech stimuli can be varied in such a way as to affect politeness judgements. This pilot study used a formant synthesiser to produce varieties of utterances of one sentence with different prosodic features: duration (which is related to tempo), fundamental frequency (which is related to pitch) and intensity (which is related to loudness). Four native speakers of Japanese were presented with these synthetic utterances, and were asked to rate them on a politeness scale. The results showed that politeness scores varied with changes in the acoustic variables, albeit with great individual differences. In other words, the subjects did use prosody for politeness judgements.

What makes research in the area of prosody very difficult is the many-to-many relationships between variables at different levels, including the physiological, articulatory and acoustic levels; a slight change of a single vocal organ could affect various acoustic variables in very complex ways, and the relationship between acoustic variables and their perceptual counterparts is by no means one-to-one. Although the physiological and articulatory variables are very important, acoustic variables are focused on in relation to Japanese politeness in the present study, because they are easier to measure and easier to manipulate at the present state of

knowledge and technology. Speech samples were collected and a number of acoustic cues were identified, such as syllable duration and fundamental frequency ( $f_0$ ). Four perceptual experiments using an acoustic feature manipulation technique examined the effects of these potential acoustic cues or features in signalling politeness in Japanese.

## **1.2. Organisation of the thesis**

The concept and relevance of Japanese politeness are the concern of Chapter 2. The concept of Japanese politeness is discussed in comparison with the Western politeness, and the reason why politeness is so important in Japanese society is explained in more detail. In Chapter 3 research on prosodic features in relation to perceived speaker variables is reviewed. The aims of this literature review are first, to obtain potentially relevant prosodic cues for signalling politeness, and second, to summarise methodological issues on speech data collection and perceptual experiments. The next three chapters are the main part of the study. The recording of speech samples is described in Chapter 4. Acoustic analyses of these samples are in Chapter 5. Three perceptual experiments based on the measurements of the samples are reported in Chapter 6. Finally, general conclusions and future work are discussed in Chapter 7.

## CHAPTER 2

### JAPANESE POLITENESS

In this chapter the concept of politeness and its importance in Japanese society is discussed. Section 2.1 compares Western politeness and Japanese politeness ('keigo'). Section 2.2 examines the status of 'keigo' at the present time and in the near future. Section 2.3 examines the recent change in emphasis of research in the 'keigo' system from strictly linguistic forms to a wider domain including behavioural and speech studies.

#### 2.1. What is politeness?

##### 2.1.1. The concept of politeness

'Politeness' is such a broad concept that it has come to be associated with a number of adjectives: good-mannered, respectful, considerate, decent, pleasant. According to Loveday (1981), the term polite covers "a whole range of notions such as sincerity, demonstration of interest, warmth, deference, social recognition, etc." (p. 71). People appear confident about the meaning of politeness in any particular situation, yet it is difficult to describe exactly what it is. In fact, Watts *et al.* (1992) in their introduction to politeness studies in language point out that "one of the oddest things about politeness research is that the term 'politeness' itself is either not explicitly defined at all or else taken to be a consequence of rational social goals ..." (p. 3).

Ide *et al.* (1992) investigated the concept of politeness from the point of view of both American speakers and Japanese speakers. They first examined various definitions used by researchers in this area in order to determine the essential meaning of the word



'politeness': "a means of minimising the risk of confrontation in discourse" (Lakoff, 1989, p. 102); "to be polite is to abide by the rules of the relationship. The speaker becomes impolite just in cases where he violates one or more of the contractual terms" (Fraser and Nolan, 1981, p. 96); "what politeness essentially consists in is a special way of treating people, saying and doing things in such a way as to take into account the other person's feelings" (Penelope Brown, 1980, p. 114). Ide *et al.* (1992) conclude that the common feature among these definitions for politeness is "the idea of appropriate language use associated with smooth communication" (p. 281). This definition is also adopted in this study.

### **2.1.2. Politeness as a strategy**

Theories of politeness have been investigated in the light of language use in social interaction (e.g., Lakoff, 1973; Brown and Levinson, 1978, 1987; Leech, 1983). Brown and Levinson's theory provides the most comprehensive account of politeness phenomena, both in verbal and non-verbal behaviours, by viewing politeness related activities as strategies. They introduce the concept of 'face', after Goffman (1967), which is "the public self-image that every member wants to claim for himself". 'Face' consists of two aspects: 'negative face' and 'positive face' (Brown and Levinson, 1987, p. 61). The 'negative face' refers to "the basic claim to territories, personal preserves, rights to non-distraction - i.e., to freedom of action and freedom from imposition" (p. 61). The positive face refers to "the positive consistent self-image or 'personality' claimed by interactants" (p. 61). They assume that members of society have a basic desire not to threaten the face of others, and a desire that their own face not be threatened by others. However, there are certain kinds of acts which intrinsically threaten 'face'. These acts are called face-threatening-acts (FTA). Two different types of politeness strategies for performing an FTA are then recognised: (1) positive politeness strategy, an 'approach-based' strategy, in which the speaker wishes to share



the addressee's wants, with an emphasis on their similarities; (2) negative politeness strategy, an 'avoidance-based' strategy, in which the speaker recognises and respects the addressee's freedom from imposition (Brown and Levinson, 1987, pp. 68-71).

The seriousness of an FTA is calculated as a linear function of various social factors with a certain weight, major factors of which are the social distance (D) of the speaker and the addressee, the relative power (P) of the speaker and the addressee and the ranking (R) of impositions in the particular culture (Brown and Levinson, 1987, pp. 74-83). So, cultures with high levels of these factors (e.g., "those lands of stand-offish creatures like the British (in the eyes of the Americans), the Japanese (in the eyes of the British)", etc.) are more likely to take FTAs as more serious than cultures with the low level factors, and thus tend to use negative politeness strategies (Brown and Levinson, 1987, p. 245).

### **2.1.3. Western politeness vs Japanese politeness**

In a broader sense, as we have seen in the definitions of politeness in Section 2.1.1, politeness is universal. The key concept of politeness is indeed an idea of appropriate language use for smooth communication in any language community. However, there seem to be some differences when it comes to the meaning and the function of politeness in different cultures. Brown and Levinson's politeness theory (1978, 1987), based on studies of three different languages and cultures (i.e., the Tamil of South India, the Tzeltal in Mexico and the English of the USA), has provided a good framework, and accounts for many politeness related phenomena quite well. However, there remains as yet no truly comprehensive theory of politeness, due to the fact that theories have been built on limited data and it is almost impossible to thoroughly investigate politeness in all the cultures in the world. This weakness has been recognised and has greatly stimulated cross-cultural studies since these politeness

theories appeared. For example, Blum-Kulka and Olshtain (1984) investigated requests and apologies in Hebrew; Hill *et al.* (1986) conducted a questionnaire survey of requests both in American and Japanese subjects; and Matsumoto (1988) and Mao (1994) re-examined the key concept of 'face' in Brown and Levinson's theory based on evidence from such oriental cultures as Japanese and Chinese.

Ide *et al.* (1992) compared the concept of politeness in the American context and Japanese context. They ask whether or not 'teinei' (which roughly corresponds to the English 'polite' in Japanese) is different from 'polite', and if it is, what the difference is. They conducted a questionnaire survey with 219 American and 282 Japanese college students. The subjects were given a grid containing descriptions of 14 interactional situations and a list of 10 adjectives ('polite', 'respectful', 'considerate', 'pleasant', 'friendly', 'appropriate', 'casual', 'conceited', 'offensive' and 'rude' for American subjects, and their equivalents for Japanese subjects). An example of the interactional situation is: 'Suppose you were an assistant professor. You made a critical comment on a student's term paper and asked him/her to rewrite a section. The student replied (A) "I'm sorry. I do see your point.." and (B) "I see. I'll give it another try...". Then the subjects were asked to encircle yes/no/NA for each adjective whether the adjective reflected their feelings. The rank-order correlation coefficient ( $r_s$ ) was calculated to assess the correlation between 'polite'/'teinei' and the other adjectives. The results showed that, in both American and Japanese subjects' responses, adjectives such as 'respectful', 'considerate', 'pleasant' and 'appropriate' had high positive correlation with 'polite' ( $r_s > 0.7$ ), and such adjectives as 'conceited', 'offensive' and 'rude' had high negative correlation ( $r_s \leq -0.7$ ), as was expected. There was, however, one significant difference found for the relation between 'polite' and 'friendly': a very high positive correlation for the American subjects ( $r_s = 0.9$ ) while a rather negative one for the Japanese subjects ( $r_s = -0.3$ ). Apparently, 'friendliness' was interpreted as 'over-familiarity', which is not polite especially in a hierarchical society, by the Japanese



subjects. Based on the results, they concluded that "studies of cross-cultural politeness cannot assume equivalence of key concepts, but must identify structural patterns of similarities and differences" (p. 293).

Hill *et al.* (1986) conducted a cross-cultural study of requests for a pen both in American English and in Japanese, with the long-term goal of comparing the system of politeness in these two very different cultures by identifying the common elements and essential differences. They identified two major aspects in a system for polite use of a language: "the necessity of speaker Discernment and the opportunities for speaker Volition" (p. 349). The word 'Discernment' is one way of translating the Japanese concept of 'wakimae', which is "fundamental to politeness in Japanese" (p. 347). Since there seems to be "no single English word [which] translates wakimae adequately" (pp. 347-348), I will use the term 'wakimae' from hence forth. The concept 'wakimae' refers to "the almost automatic observation of socially-agreed-upon rules" or, in other words, "conforming to the expected norm" (p. 348). The other concept 'volition', which is complimentary to 'wakimae', is "the aspect of politeness which allows the speaker a considerably more active choice, according to the speaker's intention, from a relatively wider range of possibilities" (p. 348). They argue that both factors, 'wakimae' and 'volition', are shared by both American and Japanese politeness systems, with a different weight of emphasis: 'Volition' is a relatively dominant factor for the American system whereas 'wakimae' is a primary factor for the Japanese system.

Matsumoto (1988) also acknowledges the importance of the 'wakimae' aspect in Japanese politeness in her examination of the universality of the notion of 'face', which is fundamental to Brown and Levinson's (1978, 1987) theory. She states that "what is of paramount concern to a Japanese is not his/her own territory [which is the key concept of Brown and Levinson's 'negative face'] but the position in relation to the others" (p. 405) and "loss of face is associated with the perception by others that one

has not comprehended and acknowledged the structure and hierarchy of the group" (p. 405). This acknowledgement of the relative position of others is extremely important in Japanese social life because of the 'vertical' nature of the social structure (Nakane, 1967, 1970), which focuses on the relationship between people with different social statuses rather than those between people with the equal status. This difference in position in society has been reflected in various social rules and norms, including language use.

The Japanese language has a very rich honorific system ('keigo'), having special words and particles with different politeness levels (e.g., Martin, 1964, 1975). For example, the Japanese word meaning 'to go' has different politeness forms: 'iku' is a plain form and 'irassharu' a respectful form. There are also particles such as 'desu' and 'masu' for expressing politeness. With the combination of these forms, there are at least three forms to express 'are you going?' in Japanese:

(a) to a close friend

iku?

'go (plain)'

(b) to an acquaintance who is slightly older than the speaker

iki-masu?

'go (plain) - politeness particle'

(c) to a person who is much older and/or higher in status than the speaker

irasshai-masu?

'go (respect) - politeness particle'

Matsumoto (1988) considers the system as a "relation-acknowledging device", saying that "taken in their broader sense, honorifics are morphological and lexical encodings of social factors in communication, such as the relationship between the interlocutors, the referents, the bystanders, the setting, etc." (p. 414). In my own



experience as a Japanese person, it seems to be more difficult to make friends with people who are much older or younger than myself in the Japanese speaking communities than in non-Japanese speaking communities. The "relation-acknowledging" aspect of the 'keigo' system appears to be one factor which could explain this. On many occasions the speaker of Japanese must select a certain level of politeness (because it is morphologically and lexically encoded, and 'neutral' level expressions do not always exist), which automatically expresses, and therefore focuses on, the vertical differences between the speaker and the addressee whether the speaker likes it or not.

The importance of the 'wakimae' factor (i.e., the nature of "conforming to the expected norm") in Japanese politeness is also supported by relatively less variation of expressions in certain situations used by the Japanese people compared with the variation found in the other language communities. A number of researchers have observed that Japanese people tend to use more conventionalised expressions in conversational exchanges (Sugito, 1981; Coulmas, 1981, p. 90; Hill *et al.*, 1986; Minami, 1987, pp. 55-56, 183-185; Matsumoto, 1988, p. 414). For example, Sugito (1981) reports that German expressions have more varieties than Japanese expressions in most of the situations studied in a survey on greeting forms, conducted at Kokuritsu Kokugo Kenkyuujo (the National Language Research Institute or NLRI for short, in Japan) during 1977 and 1981 both in Germany and in Japan. Hill *et al.* (1986) also found very high agreement on the appropriate forms for making a particular request among their 525 Japanese subjects when they were given hypothetical situations characterised by the addressee features such as occupation/status, relative age, familiarity. On the other hand, their 490 American subjects showed a more diffuse correlation between the person/situation features and the appropriate forms in the same questionnaire survey.



In summary, expressing politeness, especially appropriate use of the honorific system in Japanese society is not a strategy as Brown and Levinson's theory claims. Politeness use in the Japanese context rather seems to be a necessity: people must acknowledge and show the relative position in the social structure by means of conforming to social norms, including the appropriate use of language in a given situation. Failing to use the system appropriately could directly lead to the speaker's social embarrassment or loss of face. Therefore, Japanese speakers do not have much freedom in selection of politeness levels.

## **2.2. 'Keigo' use in the future**

As we have seen in the previous section, the emphasis in terms of use in the 'keigo' system is still much on the hierarchical relationships between the speaker and the addressees, reflecting the Japanese 'vertical' social structure. However, Japanese society has been changing from the rigid hierarchical structure to a more solidarity-based one, mainly due to the Western cultural and industrial influences since the end of the Second World War. Along with changes in society, 'appropriate' use and actual use of the 'keigo' system have been changing too. For example, my grandmother and her children used to use 'keigo' to her husband at home, while my mother and I do not use 'keigo' to my father. In fact, two surveys which were conducted at Kokuritsu Kokugo Kenkyuujo (NLRI) in 1953 and 1972, in the small town of Okazaki in Aichi Prefecture in Japan, with more than 400 people taking part, support my personal experience (NLRI, 1957, 1983). In their questionnaire surveys, they asked informants whether or not 'keigo' should be used to their senior/superior at home: in the 1953 survey, 42% of the informants answered 'yes'; however, in the 1972 survey, only 20% of the informants said 'yes'. So the tendency clearly goes to simplification in terms of 'keigo' use at home.

Does this mean that 'keigo' is disappearing? The answer seems to be 'No'. Minami (1987, pp. 197-209) considered to what extent the 'keigo' system would be in operation in the near future. He suggested three possibilities following Miyaji (1985): (1) more elaborate use, (2) more simplified use and (3) combination of (1) and (2). Minami believes the latter (3) to be the most probable development. He predicts more simplified use to people whom the speaker knows very well and more elaborate use to people with whom the speaker needs to establish and maintain good relationships, especially in business settings. He refers to the results of three surveys which focused on the forms of appropriate expressions in various situations, as evidence to support his claim (p. 201). The surveys are the two conducted at NLRI (1957, 1983) in Okazaki, and a similar survey with more than 500 residents of a big city, Sapporo in Hokkaido (Shibata *et al.*, 1980). The results show that expressions in situations where high levels of politeness were needed became more elaborated, while expressions in 'non-polite' situations became more impolite or casual.

The changes in social structure, including the changes in the Japanese industrial structure and in people's life style, seem to be gradually changing the function of 'keigo' in society. The function of showing the relative power status of the speaker and addressees remains dominant because the Japanese social structure is still a 'vertical' one despite various changes in society (Minami, 1987, pp. 114-116). However, a new function of 'keigo' seems to have emerged: maintaining adequate distance between the speaker and the addressees. This aspect of 'keigo' use seems to become more important as the life style becomes more urbanised and individualised (e.g., NLRI, 1986; Minami, 1987, pp. 204-206). So the importance of 'keigo' in Japanese society will remain the same regardless of changes in social structure and subsequent changes in function of 'keigo'.



### 2.3. Empirical studies of 'keigo'

Since 'keigo' is very important in Japanese society, but very complex in linguistic form, it has received a great deal of interest of linguists mainly from a descriptive point of view (e.g., Martin, 1964). The 'keigo' system has also received particular attention from sociolinguists due to the fact that the key concept of politeness involves adequate use of language systems in social situations. According to Ide (1986), in her review on the background of Japanese sociolinguistics, "the main interest of the field is concentrated on linguistic variety according to regions, and on speech behavior in daily life" (p. 281). Kokuritsu Kokugo Kenkyuujo (NLRI) has performed a leading role in this area since it was established in 1949 for the purpose of "doing scientific research on the Japanese Language and on the speech behavior in the daily life of the Japanese people, as well as establishing solid bases for improving the Japanese Language (Article one of the legal document establishing the Institute)" (p. 281). The size of the language surveys conducted by the institute is very large, involving more than 10 researchers consisting of linguists, sociolinguists and statisticians, together with a number of assistants, dealing with hundreds of informants. The methods of collecting data used in the surveys are questionnaires, interviews, observations in experimental settings and the recordings of speech samples in natural settings. The politeness related topics include 'keigo' and the knowledge of 'keigo' (NLRI, 1957), 'keigo' in private enterprises (NLRI, 1982), the two surveys on 'keigo' use and the knowledge of the 'keigo' system in Okazaki (NLRI, 1957, 1983) and changes in society and standards for politeness behaviours (NLRI, 1986).

Although various aspects of 'keigo' use in daily life have been studied using a number of different methods, including using recordings of the actual conversations (e.g., NLRI, 1971), the main focus has been limited to linguistic forms, which can be transcribed only in phonemic symbols, and very few researchers have focused on

speech and behavioural aspects of politeness to date. Hong (1993), who studied the prosodic aspects of politeness in Japanese speech, also acknowledges that empirical research on paralinguistic aspects of politeness has just begun in this area, although the importance and therefore, the necessity of such research has been pointed out by a number of researchers (e.g., Kindaichi, 1964; Nomoto, 1974; Minami, 1987).

Several factors seem to have hindered researchers from conducting empirical research on the aspects beyond linguistic forms: first, Japanese people have been more sensitive to linguistic forms in relation to situations than tone of voice or attitudes, because mistakes in the selection of appropriate forms can be instantly spotted by others, and they somehow manage to learn appropriate ways of speaking and appropriate attitudes reasonably well as they grow up in Japanese society; second, cross-cultural studies on these aspects of voice and attitude have only been stimulated recently by a number of misunderstandings which have taken place between learners and native speakers, as the number of learners of the language increases. For example, inadequate pronunciation and prosody of the learners are reported to be important factors for miscommunication (e.g., Otsubo, 1990; Hong, 1992); third, researchers working in the speech technology areas have shifted their interest. For example, researchers in the area of speech synthesis have become more interested in naturalness than intelligibility, which was the main focus until the mid-eighties. Attempts to make synthetic speech sound more human-like began, which naturally lead the researchers to show more interest in prosodic aspects (e.g., Murray *et al.*, 1988; Abe and Sato, 1993); finally, computers and computer software, which allow researchers to analyse and manipulate speech and behavioural data in a controlled way, have only recently become more accessible to people who do not have very strong engineering backgrounds. Thanks to these changes outlined above, the time is now ripe for more empirical research on various aspects of language use, which could not easily have been pursued before.



## 2.4. Summary

The universal aspect of politeness can be well captured by the definition "the idea of appropriate language use associated with smooth communication". However, there is a difference between Western politeness and Japanese politeness in terms of its meaning and its function in society. Two aspects of politeness have been discussed: 'wakimae' and 'volition'. 'Wakimae', which is fundamental to Japanese politeness, refers to conforming to the social norms, whereas 'volition', which is more important to Western politeness, rather emphasises more active choice of the speaker. The concept of 'wakimae' is realised as various social norms including language use. The Japanese language has a very rich honorific system, called 'keigo', which can be seen as a relation acknowledging device. Although changes in social structure from a vertical one to a flat one has been changing the function of 'keigo' in society, there is no doubt that the system still plays a very important role in every aspect of Japanese social life. The importance of 'keigo' has made Japanese linguists and sociolinguists rigorously study the system, yet the focus is limited to linguistic forms. More empirical research on various aspects of language use has just begun for various reasons which include the availability of technology and the shift of interest stimulated by cross-cultural studies.



# CHAPTER 3

## RESEARCH ON PARALINGUISTIC CUES TO SPEAKER VARIABLES

### 3.1. Introduction

This chapter discusses various issues relevant to conducting research on speech and perceived speaker variables. The term 'speaker variables' could include a wide variety of phenomena including the speaker's physical, psychological and social states. I use this term in this thesis as a general term covering the vocal and nonverbal aspects of communication. These signal the speaker's affective, attitudinal and self-representational aspects. Section 3.2 defines what features are covered by 'prosodic' and 'paralinguistic' in this thesis. Section 3.3 reviews studies which have investigated objectively measurable vocal features in relation to speaker variables in order to obtain potential acoustic cues for signalling politeness in Japanese speech. Finally, Sections 3.4 and 3.5 concern methodological issues for collecting speech data and preparing perceptual experiments.

### 3.2. What do 'prosodic' and 'paralinguistic' mean?

There can be different schemes to define 'prosodic' and 'paralinguistic' systems, and there is inconsistency in usage of these terms in the literature (e.g., Crystal, 1969, p. 177; Laver and Hutcheson, 1972, pp. 11-13).

In this thesis I use the term 'prosodic features' to refer to vocal features which have close ties with linguistic structures, such as pitch, tempo, pause and loudness, and use 'paralinguistic (vocal) features' to include both prosodic features and quality of voice.

### **3.3. Potential paralinguistic cues to speaker variables**

#### **3.3.1. Introduction**

There are studies which have investigated paralinguistic features in relation to politeness in different languages. Potentially relevant paralinguistic features for politeness are, for example, tempo in terms of articulation rate in English (e.g., Brown *et al.*, 1974), pitch level and voice quality in Tzeltal (e.g., Stross, 1977) and final pitch movement in German (Scherer *et al.*, 1984). However, people usually learn how to be polite in society (i.e., 'how to be polite' highly depends on the community's conventions and expectations), hence display rules could be very culture- and language-specific. So I first review studies which have directly focused on politeness in Japanese speech in Section 3.3.2. In Section 3.3.3 I then extend the scope of my review to studies on acoustic properties related to major paralinguistic features (i.e., pitch, tempo, loudness and voice quality) for signalling speaker variables (including politeness) in different languages. The purpose of the more general review is to identify potentially relevant acoustic cues to speaker variables.

#### **3.3.2. Potential paralinguistic cues to Japanese politeness**

There are very few studies which have focused on politeness in Japanese speech. In order to examine what variables are known to be related to politeness in Japanese speech, I now review two studies, which have investigated paralinguistic cues to Japanese politeness.

The first one is Loveday's (1981) study on the pitch level of polite utterances by Japanese and English of both sexes. Five native speakers of Japanese (3 male and 2 female) were asked to read a short written dialogue in a certain role (e.g., greeting an

acquaintance) in both Japanese and English. Both the English and Japanese dialogue (which was translated from the English version into Japanese) included such greetings as "oh hello", "bye" and "thank you". Five native speakers of British English also took part in reading the English lines for comparison.

The interesting findings were: first, the Japanese males consistently adopted a much lower level of pitch in Japanese (most of the  $f_0$  values were below 100 Hz) than the English males (most of the  $f_0$  values were well above 100 Hz); second, in contrast with the very low pitch level of the Japanese male subjects, the Japanese female subjects adopted a slightly higher pitch level in Japanese (extreme  $f_0$  values at the peak and the end of each utterance: 400 Hz - 190 Hz) compared with the English females in English (extreme  $f_0$  values at the peak and the end: 320 Hz - 110 Hz). He concludes that the pitch level is very differently used depending on the sex of the speaker in Japanese compared with English speakers. Although female voice is usually higher in pitch than males, the Japanese female speakers adopted relatively higher pitch level than the English females while the Japanese males adopted relatively lower pitch level than their English counterparts. Considering the fact that females are expected to be more polite than males in Japanese society, these findings suggest that pitch level could be a cue for signalling politeness.

The second study is Ogino and Hong (1992), which is the first major work with the specific focus on acoustic properties for signalling politeness in Japanese speech. They conducted a series of studies consisting of questionnaire surveys on Japanese people's knowledge about politeness in speech, and acoustic analyses of polite and non-polite utterances in terms of acoustic variables such as  $f_0$ , duration and intensity.

Among a series of surveys conducted by Ogino and Hong during 1989 and 1991, I introduce one survey in which they focused on what criteria native speakers of



Japanese would use to judge politeness of the speaker. A total of 223 people (97 male and 126 female, aged between 23 and 74 with the average age 45) living in Tokyo in 1991 took part in interviews by the researchers. The major findings were as follows. Firstly, it was found that native speakers of Japanese would rely on the appropriateness of the honorific forms most, followed by facial expressions, tone of voice, gaze, attitudes and appearance. Secondly, as for prosodic features, tempo and pitch movement were considered to be more important than loudness and pitch level for their judgement of politeness in speech. Finally, when the informants were asked to encircle the types of speech which they thought to be polite, on a given list, most of them selected slow, low-pitched and soft speech.

In order to investigate acoustic characteristics of polite speech, they made recordings of two sentences spoken by 12 native speakers of Japanese (6 male and 6 female) with two different speaking styles, polite and non-polite. The speakers varied in age between 30 and 54, and half of them had professional acting experiences. All the speakers except one were from Tokyo/Kanagawa areas. The sentences used are: (A) *Kokokara Ginza made donokurai kakarun deshouka* (meaning 'From here to Ginza, how long would it take?') and (B) *Moshimoshi, Tanaka-san no otaku desuka* (meaning 'Hello, is that Mr. Tanaka speaking?'). Two occurrences of 48 utterances (i.e., 2 sentences x 2 speaking styles x 12 speakers) were then recorded onto tape in random order, and presented to a total of 202 Tokyo residents (82 male and 120 female, aged between 23 to 74). The listener-judges were asked to rate each utterance on a 4-point politeness scale ranging from 1 (does not sound polite) to 4 (sounds very polite).

Based on this politeness assessment, the male and female utterances were ordered separately from the highest politeness scores to the lowest, and for each group (i.e., male and female) the upper half was categorised as 'polite' and the lower half was 'non-polite'. The following acoustic variables in each utterance were measured:



speaking time (in milliseconds or ms), pause time (in ms), maximum intensity of the first and the second phrase (in dB), maximum f0 of the first and the second phrase (in Hz), and utterance final intonation in terms of f0 direction and duration. The major findings are: (1) utterances which had longer speaking time and longer pause time were rated more polite, although pause time varied greatly depending on the speaker; (2) the maximum intensity levels were not significantly different in the polite and non-polite utterances; (3) the maximum f0 values were not significantly different in both speaking styles, except for the female utterances of one sentence; (4) the final f0 movement (i.e., direction and duration) seems to be important for politeness: a falling tone was adopted by all the polite utterances of Sentence (A) while no clear patterns were found for Sentence (B); (5) the sex of the speaker showed no significant difference in the politeness ratings.

In summary, the acoustic properties which have been investigated and could be a cue to Japanese politeness are temporal variables (e.g., speaking time, pause time and speech rate) and pitch related variables (e.g., f0 level and range). The sentence final intonation (i.e., f0 direction and duration of the utterance final positions) seem to be very influential for perceived politeness. The intensity variables were found to be non-significant to distinguish politeness levels, although 'soft' speech was considered to be polite by native speakers of Japanese. Finally, there appear to be some sex differences in using certain prosodic features for politeness (e.g., relatively higher f0 levels for Japanese women and relatively lower f0 levels for Japanese men compared with their English counterparts, in Loveday's (1981) study).

### **3.3.3. Acoustic cues to speaker variables**

This section reviews studies which have focused on the relationship between acoustic properties of speech and perceived speaker variables. The acoustic properties

underlying percepts such as pitch, tempo and loudness have received a great deal of interest from researchers in different disciplines including psychology, psychiatry, engineering sciences and linguistics. For example, psychologists and psychiatrists have been interested in the acoustic and perceptual properties of speech in relation to personality, emotion and mental states; engineering scientists, in relation to speaker identification and speech synthesis; and linguists, in relation to the significance of intonation patterns and conversation regulation. There is a large volume of literature in each area, and a comprehensive review of literature is beyond the scope of this thesis. Good reviews are provided in Kramer (1963), Crystal (1969, Chapter 2) and Scherer (1979a, 1979b, 1982). More recent reviews can be found in Frick (1985), Murray and Arnott (1993), Pittam and Scherer (1993) and Banse and Scherer (1996). The review in this section especially focuses on the acoustic properties which have been identified in the literature. This is in order to select potentially relevant acoustic cues for the signalling of politeness in Japanese speech.

### **3.3.3.1. Acoustic variables related to pitch**

Speech sounds are complex tones, consisting of many frequency components. Among them, the lowest frequency component ( $f_0$ ), which is known to be perceived as pitch by listeners, has been extensively investigated.

Mean and median of  $f_0$  have been most commonly used for indicating the pitch level, whereas standard deviation (SD) or coefficient of variation (SD divided by mean), the difference between the peak (highest or 90 percentile point) and the floor (the lowest or 10 percentile point) of the  $f_0$  contours have been used for variability or range. Bezooijen (1984, Chapter 5) examined the relationship between acoustic measurements and perceptual ratings by calculating the product-moment correlation coefficient ( $r$ ) between measured variables (e.g., mean  $f_0$ ) and six listener-judges'



rating scores on the perceptual counterparts of the variables measured (e.g., pitch level), using 160 emotional expressions. The results show that mean and median are very efficient indicators for the perceived pitch level ( $r's \geq 0.8$ ); SD, coefficient of variation and the difference between the 90th and 10th percentile of the log-converted  $f_0$  distribution are satisfactory predictors for the perceived pitch range ( $r's \geq 0.6$ ) with the coefficient of variation being the best among the three.

Peak and floor values are also examined in relation to discourse functions (Menn and Boyce, 1982), and initial and end points in terms of social interactions (Brazil *et al.*, 1980). The rate of change or steepness has also been used to examine the speed of  $f_0$  movement (e.g., Fairbanks, 1940; Ross *et al.*, 1986; Henton, 1995).

The shape of  $f_0$  contour is apparently very important for signalling speaker variables. However, since no two  $f_0$  movements in natural utterances can be exactly the same, it is vital to separate relevant features from irrelevant features of the  $f_0$  contour in order to use this factor for research. Various attempts have been made. For example, 't Hart *et al.* (1990, pp. 72-74) propose four dimensions to describe pitch movement of Dutch: (1) direction (rise or fall); (2) timing with regard to syllable boundaries (early, late or very late); (3) rate of change (fast or slow); and (4) size (full or half). Other variables or indices used for describing or comparing  $f_0$  contours are: the average number of change in direction per second during phonation (Fairbanks, 1940), patterns of the slopes of regression lines (Takefuta, 1975),  $f_0$  fall-rise patterns (Cosmides, 1983) and similarity index calculating the difference between a time-adjusted  $f_0$  contour and a neutral  $f_0$  contour of the same utterance (Ross *et al.*, 1986).

Among features concerning the shape of  $f_0$  contours, the final  $f_0$  movement has received particular attention in signalling politeness in Japanese (Ogino and Hong, 1992). The importance of the final part as an information carrier is seen too in German

in expressing speaker variables including politeness (Scherer *et al.*, 1984), in Sichuanese Mandarin in expressing affect (Chang, 1958), in distinguishing contrasted intonation in American English (Takefuta, 1972) and controlling and structuring the flow of discourse in English (Brazil *et al.*, 1980). So the final part has importance beyond Japanese.

Various implications of these f0 features in terms of attitudes, emotions and discourse structures have been found in the literature. I introduce some of them which are closely related to politeness. For example, Ohala (1984, p. 2) states that although the universality of affective use of f0 is not entirely conclusive, "it seems safe to conclude that such 'social' messages as deference, politeness, submission, lack of confidence, are signalled by high and/or rising F0 whereas assertiveness, authority, aggression, confidence, threat, are conveyed by low and/or falling F0" by citing Bolinger's (1964) study. This is supported by a number of findings: the high-pitched voice adopted by Japanese women for politeness (Loveday, 1981; Ogino and Hong, 1992); falsetto for deference by Tzeltal speakers in Mexico (Stross, 1977; Brown and Levinson, 1987, p. 267); and the high-pitched voice's association with hesitation (Ladd, 1980, p. 105) and connectedness (McLemore, 1992).

However, some studies show contradictory findings. Loveday (1981) reports that Japanese men adopted very low f0 levels for polite expressions, and Scherer's (1979b) findings show that high f0 levels were associated with competence, dominance and assertiveness for American males, while the high f0 levels received higher ratings on the axes of discipline and dependability for the male German and female American subjects.



### 3.3.3.2. Temporal variables

There are several temporal variables which have been investigated in relation to speaker variables in the literature. Among them, the following are the variables which have been identified as potentially relevant: speech rate, variables related to pauses, and duration of the total and some parts of utterances. In addition to these variables, the potential importance of rhythm or micro temporal structure has been acknowledged by researchers (e.g., Miller *et al.*, 1984; Brown and Bradshaw, 1985), but few studies have focused on this, perhaps reflecting the difficulty of defining rhythmicity.

Speech rate (usually measured in syllables per second) has been widely used as a major acoustic correlate of tempo. There are two types of rate measures: one is inclusive of pauses, and the other is exclusive of pauses (which is also called articulation rate). The correlation between these two rate measures and perceived tempo is reported to be very high; 0.77 for the speech rate with pauses and 0.85 for the speech rate without pauses, based on Bezooijen (1984, p.64).

Because of the high correlation with perceived tempo, the speech rate has been most commonly used in both acoustic analyses and cue manipulation experiments. The effects of speech rate on ratings of speaker variables are reported to be very consistent, and much greater than the other aspects of speech (e.g., f0 level, f0 variation, intensity) (e.g., Brown *et al.*, 1974; Scherer and Oshinsky, 1977), although this largely depends on the specific acoustic variables included in the study and on the extremity of the values used for the variables studied. Brown *et al.* (1974) and Smith *et al.* (1975) conducted experiments on the relationship between articulation rate of utterances and speaker variables of competence (e.g., active, intelligent, confident) and benevolence (e.g., kind, polite, just). Major findings are: (1) fast speech was generally associated

with high competence, whereas a relatively slower or normal rate of speech tended to receive a higher rating on benevolence axes; (2) slow speech decreased both competence ratings and benevolence ratings. In other words, the highest benevolence ratings were given to voices in the middle range of measured rate of utterances. This is supported by the results of Experiment 1-B (reported in Section 6.2). However, Brown, Giles and Thakerar (1985) found a linear relationship between articulation rate and benevolence variables (i.e., the slowest utterance had the highest benevolence ratings). This contradictory result may be due to differences in experimental design, subjects and articulation rate values used in these experiments.

The importance of pauses (silence or filled) has been recognised. Pauses are generally associated with the fluency aspect of speech, and are known to influence perceived tempo; speech containing a high proportion of pauses tends to be evaluated slower than otherwise (Sugitou, 1986). Therefore, variables related to pauses can be important in signalling politeness in terms of hesitation and indirectness. The pause-related variables are: the number of pauses per utterance and duration of pauses, in relation to hesitation and emotion (e.g., Scherer, 1979b, pp. 160-168; Cosmides, 1983; Ogino and Hong, 1992), and the ratio of pause duration to speech duration in terms of fluency and affect (e.g., Scherer, 1979b, p. 161; Ross *et al.*, 1986).

Major variables in utterance duration are total speaking time and duration of the final part of utterances. Considering the importance of the final part of utterances as an information carrier, as we have seen in the previous section, the duration of the final part can be a powerful indicator for politeness, which is, in fact, supported by the results of Experiment 1-A (Section 6.1). It is known that utterances with longer speaking time were perceived as more polite than those with shorter speaking time in Japanese speech (Ogino and Hong, 1992), and the importance of duration of the final



part of utterances has been acknowledged in distinguishing comfort and discomfort in Japanese (Imaizumi *et al.*, 1994).

### 3.3.3.3. Acoustic variables related to loudness

Several variables associated with amplitude or intensity have been measured in relation to emotions: amplitude level, variability, rate of change (Ross *et al.*, 1986), amplitude fall-rise patterns (Cosmides, 1983). Peak intensity values in phrases are also measured in relation to Japanese politeness (Ogino and Hong, 1992).

Although perceived loudness is thought to be important, the acoustic variables directly related to intensity or amplitude are not found to be significant for signalling speaker variables. No strong significant correlations between objectively measured intensity and emotions nor personality attributes have been reported neither in externalisation studies (e.g., Cosmides, 1983), nor in cue manipulation studies (e.g., Lieberman and Michaels, 1962; Scherer and Oshinsky, 1977), although it may only show that there are some methodological problems in elicitation methods, cue manipulation techniques or rating sessions. Frick (1985) argues in his review that the relevant feature may be vocal effort, and not intensity. Brandt (1972) suggests that a likely cue to perceived vocal effort is, in fact,  $f_0$ .

The distinction between intensity and perceived loudness appears to be suggested by comparing findings in the literature. For example, subjective ratings of loudness are found to have a significant negative correlation with dominance (Mallory and Miller, 1958) but Scherer (1979b, p. 158) did not find intensity to be significant for dominance; Ogino and Hong's survey (1992) on tone of voice for politeness in Japanese shows that Japanese people consider 'weak/soft' voice is a cue to politeness but they could not find any significant correlation between peak intensity and



politeness. An important factor appears to be vocal effort or tension and relevant acoustic variables may be found in the areas of articulation or voice quality. The acoustic variables related to tension are discussed in the next section.

#### **3.3.3.4. Acoustic variables related to voice quality and articulation**

Voice quality and articulation are apparently important just as pitch, tempo and loudness for expressing speaker variables. Various attempts have been made to assess the effects of voice quality and articulation. Auditory assessment was mostly used to study voice quality in relation to introversion, dominance, sociability and emotional stability (see Scherer, 1979b). This approach has been criticised for the impressionistic labels used, and the search for objectively measurable variables has been undertaken.

Differences in voice quality can be more easily described from the production point of view (e.g., a voice produced by a raised larynx) than from the point of view of acoustic characteristics of the speech signals, and thus the voice quality has mainly been described using articulatory terms than acoustic terms. Two factors are considered to be important for distinguishing one voice from another: the anatomical and physical characteristics of a speaker's vocal organs and the muscular adjustments of these vocal organs (e.g., Laver, 1968). Since the physical aspects (e.g., the shape and the size of vocal organs) are usually well beyond the speaker's control, the aspects of muscular adjustments have been a focus in studies on speaker variables.

There are various factors for the muscular adjustments or settings. Laver (1980) categorised them into three types: the settings of the vocal tract (Supralaryngeal settings), the phonatory mechanisms of the larynx (Phonatory settings) and the settings of overall degrees of muscular tension throughout the vocal system (Tension settings). The supralaryngeal settings have three types for modifying the shape of the vocal tract:

longitudinal modifications (e.g., by raised or lowered larynx), latitudinal modifications (e.g., by the movement of the lips and the tongue) and velopharyngeal modifications (e.g., by opening or closing the entrance to the nasal chamber). The phonation types include breathiness, whisper, creak, falsetto and harshness in relation to the 'neutral' voice. The tension settings distinguish 'tense' and 'lax' voice. The tense voice is also labelled as 'metallic', 'clear' or 'sharp', and the lax voice, 'muffled', 'dull' or 'soft'.

Among various types of voice mentioned above, the tense/lax voice seems to be related to perceived loudness (e.g., Laver, 1980, p. 148). Since loudness is certainly a very important factor for politeness, as we have seen in Ogino and Hong's (1992) survey, I particularly focus on the aspect of tension and related phonation types in this section. One of the important acoustic variables for tension seems to be the relative amount of energy in the upper harmonics (e.g., Dusen, 1941; Frøkjær-Jensen and Prytz, 1976, p. 3). Other variables which seem to be related to tension and vocal effort are: the shape of glottal waveform (e.g., skewness of glottal pulse) and the spectral energy distribution (e.g., Monsen and Engebretson, 1977; Bezooijen, 1984, pp. 61-63; Childers and Lee, 1991).

The tense voice is often associated with harshness and the lax voice tends to co-occur breathiness. The harshness is caused by irregularities in the activities of the vocal cords (e.g., Borden and Harris, 1984, pp. 87-88). Therefore, perturbation in  $f_0$  or in amplitude have been used for an indicator for harshness (e.g., Fairbanks, 1940; Wendahl, 1963, 1964; Bezooijen, 1984, pp. 60-61; Klasmeyer and Sendlmeier, 1995).

The breathiness is basically achieved by failing to adduct the vocal cords sufficiently enough for full voice (e.g., Borden and Harris, 1984, pp. 87-88). The potential acoustic variables are the damping characteristics of the wave (e.g., Laver, 1980, p. 127; Childers and Lee, 1991), the amplitude difference between the lowest harmonic



( $f_0$ ) and the higher harmonics (Bickley, 1982; Henton and Bladon, 1985) and the ratio of the  $i$ -th harmonic amplitude to the  $i$ -th interharmonic noise (the noise-to-harmonic ratio, or NHR for short) (Childers and Lee, 1991; Krom, 1994; Klasmeier and Sendlmeier, 1995). Among these measures, the NHR appears to be a very efficient indicator for breathiness (e.g., Childers and Lee, 1991; Krom, 1994).

Articulation (e.g., precise or careless) is intuitively important for expressing politeness because of a link with carefulness. Acoustic characteristics related to the articulation type are formant trajectories (Scherer, 1982, pp. 162-163) for vowel quality, voice onset time of stop consonants (Fairbanks, 1940) and intensity of release for stop consonants (Williams and Stevens, 1972) for consonant quality. However, what measures are to be used as good indicators for the way of articulation is not still very clear at the present time.

### **3.4. Methodological issues regarding speech data collection**

This section discusses three factors which must be considered carefully for speech data collection: Elicitation methods, message content and speakers.

#### **3.4.1. Elicitation methods**

There are basically two different ways of collecting speech samples: one is 'field' recordings in natural settings and the other is laboratory recordings. There is of course a trade-off here between realism on the one hand and achieving a high degree of control on the other. Although the artificiality of the laboratory recordings could be inadequate to study ordinary speech, there are usually too many potentially relevant situational factors involved in the field recordings, as has been noted by a number of researchers (e.g., Edelsky, 1979; Loveday, 1981; Geluykens and Swerts, 1992; Ogino



and Hong, 1992). So compromise approaches, which attempt to achieve natural simulations under laboratory conditions, may be "the only realistic possibility for the research" (Scherer, 1979a, p. 509).

The simplest way to obtain simulated portrayals is to ask speakers to speak or read test passages in certain ways (e.g., politely or angrily). Descriptions of scenes to help speakers simulate certain feelings are sometimes provided in the recording sessions. However, this method tends to induce theatrical exaggeration (e.g., Cosmides, 1983).

Another method to obtain natural simulations is to use a role-play method, in which speakers are given scenarios and asked to play their roles (e.g., Scherer and Scherer, 1980; Hong, 1993). Although the role-play method may be more adequate in terms of suppressing theatrical exaggeration, there is the danger of speakers' resorting to stereotyped representations, which might not occur in natural settings (e.g., Kramer, 1963). The essential question here is whether or not these stereotypes could be 'natural' enough for studying speaker variables. Williams and Stevens (1972) address this question by comparing the real commentary of the Hindenburg disaster and an actor's simulation of it. It was found that the median  $f_0$  and  $f_0$  range both increased dramatically for the announcer and the actor after the crash, but the changes in the actor's simulation was much greater. So there is a possibility of exaggeration in any actor simulations, although this particular result may have been due to the individual difference between the announcer and the actor. The important point is, however, that the pattern of changes (i.e., the median  $f_0$  and  $f_0$  range increased) was similar for the real commentary and the simulation. Therefore, using simulations (by either experienced or inexperienced speakers) can be justifiable to a certain extent so long as simulations are not over exaggerated. Overemphasis is, in fact, reported to reduce decoding accuracy (Wallbott and Scherer, 1986).

### 3.4.2. Selection of test passages

Selection of test passages is important for both speech data collection and listening material preparation, because of potential interactions between message content and both production and perception of speech. Various attempts have been made to minimise or eliminate the content effects: using meaningless content, using constant content and masking content (e.g., Kramer, 1963). The meaningless content methods use materials which do not have any communicative meaning themselves. The 'content-free' materials used in 'affect' studies are isolated words such as 'ah' (e.g., Skinner, 1935), letters of alphabets (e.g., Davitz and Davitz, 1959), nonsense syllable sequences (e.g., Uldall, 1964) and tone sequences (Scherer and Oshinsky, 1977). The constant content methods use phrases which are neutral or ambiguous in the sense that no specific feelings or attitudes are attached to them (e.g., Fairbanks and Pronovost, 1939). For example, 'my father died' primarily implies sadness or distress while 'this is a pen' does not have any special feelings attached to it. The masking content methods use natural utterances, but ignore content by distracting listeners' attention from verbal content (e.g., Brody, 1943), or eliminate verbal content electronically, for example, by low-pass filtering (e.g., Starkweather, 1956).

Since politeness is usually closely associated with appropriateness in a specific situation, it is difficult to separate it from verbal content. In some situations (e.g., the verbal content cannot be heard clearly) people may still distinguish polite utterances from non-polite utterances. However, since the aim of the present study was to observe changes in politeness judgements assessed by rating scores, it was considered that naturalness of stimuli was essential, but content-free or electronically content-masked materials were not suitable. Therefore, the constant content method was used for the recording and preparation of stimuli.

### **3.4.3. Speaker variability**

Another important factor for speech data collection is speakers. The variability of speakers comes from various factors, including sex, age, accent, educational and social backgrounds and personality. In fact, the encoding abilities of individual speakers are reported to vary a great deal (e.g., Wallbott and Scherer, 1986). So it is very important to use more than one speaker to avoid any misleading generalisations of findings based on particular samples.

The sex of speakers is of a particular importance in studying Japanese speech, because the Japanese language has distinctive words and expressions for men and women, together with different social expectations for both sexes (e.g., Women are expected to be more polite than men), and there is also a difference in how to use prosody between male and female speakers, as we have seen in Section 3.3.2. Unfortunately, female voices are more difficult to analyse acoustically (e.g., extraction of  $f_0$  and formant structures) and synthesise, mainly due to high pitch. Since this study used both acoustic analysis and synthesis, male voices were used throughout. Another reason why male voices were focused on is that male speakers are reported to have a wider range of expressions in terms of politeness levels than female speakers (e.g., Minami, 1987, pp. 147-156).

### **3.5. Methodological issues regarding perceptual experiments**

This section concerns techniques and problems associated with acoustic cue manipulation and rating tasks. I start with the issues regarding acoustic cue manipulation.



### **3.5.1. Acoustic cue manipulation**

In order to select techniques for cue manipulation for preparing stimuli for perceptual experiments, the following two factors have to be considered: the speech quality (i.e., whether the stimuli sound natural enough or not), and the controllability of variables (i.e., which variables to manipulate and which variables to keep unchanged). Section 3.5.1.1 reviews currently available techniques for speech synthesis and acoustic variable manipulation. Section 3.5.1.2 discusses considerations on selection of these techniques.

#### **3.5.1.1. Synthesis/resynthesis techniques**

There are currently four methods to synthesise/resynthesise speech signals: (1) Articulatory synthesis, (2) Formant synthesis, (3) Linear Prediction (LP) analysis-synthesis and (4) Pitch Synchronous Overlap and Add (PSOLA) waveform concatenation. The first three techniques construct speech signals while the PSOLA technique only manipulates pre-recorded speech signals. The PSOLA and LP techniques are commonly used as a tool for modifying prosody of a given speech signal. Such acoustic variables as  $f_0$  and duration can be altered without much degradation in speech quality by using these techniques, especially the PSOLA technique.

##### **3.5.1.1.1. Articulatory synthesis**

Articulatory synthesisers are built by modelling human speech production systems in terms of the positions and the movement of articulators. Changes in the shape of the vocal tract are described as the movement of these articulators towards target positions for each phoneme. The first electric models were developed in the mid-fifties

(Stevens *et al.*, 1953; Rosen, 1958), but this approach has not yet produced good quality synthesisers compared with synthesisers achieved by other approaches.

Although articulatory synthesis seems to have a great potential for obtaining natural speech by machine, the difficulties lie with modelling the articulators' dynamics accurately mainly due to lack of data (Klatt, 1987).

#### **3.5.1.1.2. Formant synthesis**

This model is based on an acoustic theory of how speech is produced, which is called the source-filter theory. Speech can be seen as the outcome of the vocal tract's response to sound sources. The source-filter theory assumes the speech production system can be reasonably accurately described by these two factors: sound source (excitation signal) and the response characteristics of the vocal tract (the vocal tract transfer function), functioning as a linear filter.

Formant synthesisers are constructed by modelling the vocal tract transfer function with a set of resonances which model the formants of natural speech. This approach has been most successful in terms of speech quality. The first dynamically controlled formant synthesisers were developed in the mid-fifties (e.g., Lawrence, 1953). Since then, modern formant synthesisers (for example, Klattalk) have been greatly improved and have reached such a level that many male voices can be imitated nearly perfectly (Klatt, 1987). Since this formant synthesis technique allows  $f_0$ , duration and certain properties of voice quality (e.g., male/female voice) to be altered, it could be used for acoustic cue manipulation of natural speech. However, the difficulties of extracting formant structures from the natural speech have prevented researchers from using this technique as a tool for prosody modification.

### 3.5.1.1.3. Linear Prediction (LP) analysis-synthesis technique

The concept of linear prediction is not new, dating back at least to the late 1940s, and has been applied to speech data (e.g., Itakura and Saito, 1968; Atal and Hanauer, 1971). The LP model applied to speech is also based on the source-filter theory, which is briefly described in the previous section. This technique represents the speech waveform in terms of time-varying parameters related to the vocal tract transfer function and the excitation source. The basic assumption is that the current speech sample can be predicted as a linear combination of the previous samples. The predictor coefficients are determined by minimising the mean-squared error between the actual and the predicted speech samples. These coefficients, together with other parameters such as  $f_0$ , a voiced/unvoiced flag and energy level of each sample, are then used for reconstruction of the speech signal. Reasonably good results are said to be achieved relatively easily.

Since  $f_0$  is an explicit parameter for this model,  $f_0$  values can be altered easily by changing this parameter prior to reconstruction of the synthetic signal. Duration manipulations can also be achieved by changing the update rate of estimated parameters. Hence, the LP technique has been extensively used as a tool for modifying these acoustic variables. The problem is a rather limited speech quality. For example, the original vowel quality (i.e., formant frequencies and bandwidths) cannot often be realised accurately when speech is resynthesised at a different  $f_0$  from the original value (Klatt, 1987); synthetic signals tend to suffer from a fuzzy noise due to an oversimplified excitation source (Moulines and Charpentier, 1990); certain types of sounds (i.e., nasalised sounds and plosives) cannot be synthesised well due to the simplified model of the vocal tract response characteristics.



#### **3.5.1.1.4. Pitch Synchronous Overlap and Add (PSOLA) technique**

The PSOLA technique was developed at CNET in France in the mid-eighties, in the process of developing a synthesiser for French by smoothly concatenating synthesis unit signals (Charpentier and Stella, 1986). This technique provides a superb way to modify prosody and concatenate waveforms, which were traditionally performed by the LP technique. There are mainly two methods to perform the PSOLA technique in terms of the domain in which waveform modifications are performed: Time Domain (TD) PSOLA and Frequency Domain (FD) PSOLA. The TD-PSOLA is more computationally efficient than the FD-PSOLA. Although the TD-PSOLA has some problems in speech quality, which can be overcome by the FD-PSOLA, since the speech quality is generally very good, the TD-PSOLA was used for preparing the stimuli in the perceptual experiments of the present study (Chapter 6).

With the TD-PSOLA technique, wave manipulations are achieved by decomposing the speech signal into overlapping pitch synchronous short-term (ST) signals, modifying them appropriately by altering the mapping of these signals on the time axis, and recombining them by overlap-add algorithms in such a way that the output waveform realises the target  $f_0$  and duration. Duration modifications can be performed simply by repeating or deleting some of these ST signals when they are mapped on the time axis. The great advantage of this technique over the LP technique is good speech quality. This is mainly because the PSOLA technique does not use a parametric representation of the speech signal as the LP technique does. Although this parametric representation provides efficient ways for transmission and storage of speech data, it causes degraded speech quality of the synthesised signal. Some acoustical distortions (e.g., a tonal noise when unvoiced segments are stretched by more than 200%) may take place in the TD-PSOLA scheme, however, they are generally negligible if moderate modification is performed.

### 3.5.1.2. Considerations on selection of techniques

There are practically two approaches to obtain cue-manipulated stimuli: (1) to manipulate certain acoustic variables of natural utterances by using a wave manipulation technique (e. g., the PSOLA technique) and (2) to use the output of speech synthesisers. I discuss the advantages and disadvantages of these two approaches in terms of speech quality and the controllability of variables.

The great advantage of Approach (1) is obviously the naturalness in speech quality. Very natural output can be obtained with moderate changes of acoustic properties of the speech signal performed by a good analysis-resynthesis technique (e.g., the PSOLA technique), although the variables which can be manipulated are usually limited to  $f_0$  and duration. This approach also enables researchers to investigate the speaker effect by using utterances spoken by different speakers. This, however, can be a problem when target sentences have to be changed: new recording is necessary every time target sentences are changed, since wave manipulation techniques do not make speech signals themselves. Furthermore, if experiments extend over a long period of time, the original voice used in the early experiments may not be available.

Approach (2), on the other hand, has difficulties in obtaining sufficiently natural-sounding output suitable for perceptual experiments on speaker variables. Human speech production is a complex process and all the important aspects of this process cannot easily be captured well enough to make satisfactory models and algorithms for constructing speech synthesisers. Although some synthesisers, for example, formant synthesisers, are reported to be capable of imitating male voices nearly perfectly, producing a convincing female or child's voice seems to be still a problem (Klatt, 1987). The advantage of using synthetic speech over using acoustically manipulated natural speech is that the voice quality can be kept reasonably constant for different



sentences throughout a long series of experiments, since a synthesiser produces speech signals either by certain algorithms and sound sources, or by concatenating pre-recorded synthesis unit signals. Human speakers, on the other hand, cannot control voice quality: even when the same speaker speaks the same sentence the voice cannot be exactly the same. Another advantage of using synthesisers would be that certain properties of voice quality can be changed systematically by using formant synthesisers (e.g., Klattalk), although acoustic variables responsible for changes in voice quality are not well understood at the present time.

As we have seen so far, each technique has its own advantages and disadvantages. Therefore the technique has to be selected carefully according to the purpose of the experiments. The factor of naturalness of the stimuli is especially important for rating tasks. For example, people cannot rate politeness, or might change their judgement criteria, if the stimulus utterances sound very unnatural. The effect of naturalness on politeness judgement is addressed later in Experiment 3 (Section 6.3). Since the main experiments reported in Chapter 6 needed very natural stimuli, and acoustic variables studied were  $f_0$  and duration, the acoustic manipulation of natural speech with the TD-PSOLA technique was used. A pilot study on the effects of prosody of synthetic speech (Appendix 1) used SYNCON, a formant synthesiser on a BBC microcomputer (Holmes, 1986).

### **3.5.1.3. Ecological validity problems**

The previous section focused on naturalness in terms of speech quality. This section discusses naturalness in terms of ecological validity. Computer cue manipulation allows experimenters great control of major acoustic variables including  $f_0$  and duration, but at the same time, this great flexibility could raise an ecological validity question: whether or not the manipulated speech remains natural or realistic.



There are two issues here: normal range of the manipulated variables and covariation between variables. Manipulations could be too extreme to be natural or the interaction between the changed variables and other variables might result in an unexpected effect on human perception. I first discuss the normal range question: are the values used in cue manipulation within or beyond a certain normal range of individual speakers or human speakers in general?

Brown *et al.* (1974), in their pioneering work which investigated the effects of synthetically manipulated speech rate, mean  $f_0$  and  $f_0$  variation on ratings of personality, changed these acoustic variables of the original utterances quite substantially for their stimuli: the change rates were from 0.7 to 1.8 for mean  $f_0$ , 0.2 to 1.8 for  $f_0$  variation and 0.5 to 1.5 for speech rate. Scherer (1979b, pp. 185-186) raises a question about the normality of these values. For example, Terango (1966) found that the male voices judged to be effeminate were considerably lower than the typical female pitch: male voices with a median  $f_0$  of 127 Hz were judged to be rather feminine while 100 Hz voices were perceived as masculine. So if synthetically changed  $f_0$  is beyond this threshold, it would change the category of voice from a normal male voice to a strange male/female voice, which is very likely to affect listeners' judgement on personality characteristics. As for the normal range of changes in speech rate, Daniloff and Hammarberg (1974) report that their speakers could not increase their speech rate by more than 30%. In order to avoid or minimise the risk of using extreme values Brown *et al.* (1974) suggest that realism ratings should be included in perceptual experiments.

The problem of covariation of variables is more complex. In fact, it would be extremely rare that only one or two variables are changed with the others kept constant in real speech, (which researchers usually do to focus on the effects of certain

variables), because human speakers produce speech by moving vocal organs, not directly producing or modifying digital signals like most computer techniques do. For example, in order to investigate the effects of speech rate on perceived speaker variables, researchers have used computer rate manipulation, which usually linearly compresses or expands each segment of the utterances. However, this rarely or never takes place in real speech. Suppose people try to speed up their speech. Some people maintain the articulation which is used in slower speech by, for example, moving their tongue faster with the target position of the tongue unchanged. However, some people change articulation, and move their tongue slower with an easier target position.

People seem to realise tempo changes in different ways, each of which results in many changes on various acoustic variables besides segmental duration: articulation changes such as elision and assimilation, pause insertion and changes in spectral characteristics (Campbell, 1992, p. 198). Even if we focus solely on changes in segmental duration, vowels are more flexible in durational change than consonants, and the position of the phoneme seems to affect the changes caused by rate change. Bell-Berti *et al.* (1995) report that slowing down speech resulted in longer durations of utterance-initial vowels, and vowels in sentence-final positions tended not to be lengthened at slow rates, and therefore had relatively shorter durations.

Manipulations of  $f_0$  could also introduce unexpected artefacts in vowel quality and voice quality. Again, human speakers are unable to raise or lower pitch without changing other spectral properties (e.g., formant structure and energy distribution). A study was conducted to examine potential effects of computer manipulation of  $f_0$  on voice quality by comparing human manipulation of long vowels ('ae' and 'ah') and computer manipulation with the PSOLA technique of the same vowels produced by a phonetician. The details are reported in Section 6.1.3. The spectral analysis showed that the positions of the first three formants, which are mainly responsible for distinguishing vowel sounds, were well preserved both in the human and computer



manipulations, whereas the upper formants were slightly changed in both manipulations in slightly different ways. The auditory assessment confirmed that the computer manipulated versions tended to preserve the voice quality of the original vowels better than the versions manipulated by the same speaker of the original vowels. The human speaker indeed changed voice quality along with  $f_0$  change, but in a not easily predictable manner.

So far, we have seen complex covariation between variables taking place in real human speech. Changing more than one variable, however, is not a problem, if the process is a systematic one. The real problem is that these changes are not necessarily systematic nor consistent across different speakers. In spite of these potential problems concerning ecological validity caused by current computer manipulation techniques, an encouraging finding for using computer manipulation comes from Bruce Brown's (1980) experiment on a comparison between the effects of human alterations and computer manipulations on speech rate. He asked six subjects to rate human manipulated versions and computer manipulated versions on the benevolence axis (which includes kindness, politeness, etc.) and the competence axis (which includes intelligence, confidence, etc.). The results showed high correspondence between the human and computer manipulations, with the main difference being uniformly higher benevolence ratings for the human manipulated stimuli.

Comparing the advantages and disadvantages of using computer cue manipulations discussed so far, there is still a great advantage of using computer manipulation over human manipulation in terms of achieving a high degree of control of variables. It is especially so when experiments are designed to obtain a general knowledge about the effects of variables studied, provided that the synthetically manipulated stimuli sound reasonably natural or realistic.



### 3.5.2. Rating tasks

The problem with studying human perception is that there is no direct or objective way to assess perception. A neurological approach (e.g., electronic measurements of neuronal activities in the brain) could be one solution, but the relationships between these measurements and human perception are not well understood, and would be far from simple. Using subjective rating seems then to be the only possible way to assess perception at the present time. Unfortunately, human rating is subjective in the sense that ratings vary depending on various factors including how subjects are instructed, what stimuli are presented and what kind of person the subject is. These factors are discussed in the following sections. Finally, since ratings can vary depending on these factors, it is very important to assess the reliability of listener-judges' scores, which is the topic of the final sub-section (3.5.2.5).

#### 3.5.2.1. Rating methods and stimulus presentation

The method used to instruct listener-judges in rating tasks is a very important factor. Ide *et al.* (1986), in their study on politeness in American and Japanese societies, used both direct and indirect questions and compared the results. They asked subjects (1) to rate politeness levels for given expressions on a 5-point scale; (2) to rate politeness levels which they would use for given categories of addressees (e.g., 'professor', 'a middle-aged man wearing jeans'), using the same politeness scale as that used in (1); and (3) to write what expressions they would use for the addressee categories given in (2) (pp. 231-253). They named politeness scores obtained by (1) and (2), the scores of 'politeness-in-concept' (PIC for short), and scores obtained by (3), the scores of 'politeness-in-use' (PIU for short) (pp. 215-218). They report that the PIC scores were distributed in a continuum while the PIU scores were bipolar, and these two scores obtained by the Japanese subjects were more highly correlated than

those by the American subjects. Although the difference between these scores concerning 'concept' and 'use' is not entirely clear, as far as they are highly correlated, using PIC as a politeness index does not appear to cause any serious problem. Therefore, a rating scale method using a scale of politeness (i.e., asking subjects to rate the politeness level of utterances on a given scale) was used for most of the listening tests reported in this thesis.

Another method adopted in the experiment on naturalness (Section 6.3) is a two alternative forced-choice procedure (or paired comparison), in which listener-judges are presented with two utterances in succession and asked to judge which utterance has a higher degree of politeness/naturalness (Watkins and Makin, 1994). This method has both advantages and disadvantages compared with the rating scale method. The most serious disadvantage is the artificiality of judgements: people do not usually assess speaker variables in this way. Since subjects are forced to make a decision even if both utterances sound nearly the same in terms of the speaker variables studied, they may have to resort to noticeable differences (e.g., naturalness) which might not be the part of the judgement of the speaker variables in natural conditions. Another disadvantage is that the scores obtained by this method are ordinal, not interval, which can restrict the selection of statistical methods. The advantage is, however, that it is much easier to make judgements compared with the rating scale method, because subjects have a reference in each trial. Despite the disadvantages mentioned above, this may be the best possible method for rating naturalness, to which subjects tend to lose their sensitivity quickly when they hear similar-sounding stimuli many times.

In order to study the effects of manipulated variables, researchers generally have to use utterances of the same content, each of which is slightly different at a certain acoustic variable level. Using a large number of repetitions of similar-sounding stimuli is, in fact, a serious problem for assessing perception, because this introduces judges'



fatigue, and insensitivity to the speaker variables under investigation. So it is desirable to assess whether the scores obtained in these rating tasks are reliable or not, but unfortunately not very many studies have seriously examined the reliability of judgement scores. This point is discussed in Section 3.5.2.5.

The order of stimulus presentation is another factor which could affect judges' ratings. The presentation order could introduce a bias caused by, for example, a contrast effect: the same utterance may sound more polite than it really is, if the previous stimulus was an extremely impolite one. So randomising the order of stimuli is necessary to minimise these order effects.

### 3.5.2.2. Stimuli

There are two factors which can influence subjects' responses, concerning stimuli of the listening tests: the speaker (or 'voice') of the stimuli and the content of the test passages used. The main concern about the voice effects is whether or not the voices used in experiments are representative of ordinary speakers. For example, Brown *et al.* (1974) mentions a potential problem (ceiling effects) of using voices which already had a high level of speaker variables studied: the ceiling effect may obscure the effects of studied variables. Therefore, it is important to assess the voices which are to be used for perceptual experiments in terms of speaker variables under investigation.

The content effects on rating tasks have been also recognised (e.g., Apple *et al.*, 1979; Ladd *et al.*, 1985; Geluykens, 1987). Geluykens (1987) showed substantial content effects even on judgements of the intonation type (i.e., question or statement), which are considered to be rather 'categorical', as opposed to 'gradient' or 'more-or-less'. He substituted different pronouns ('I', 'you' and 'he') in such an utterance as '... not feeling very well'. Since 'you feel ill' is more likely to be interpreted as a question



compared with 'I' or 'he', 'you' has a question-prone bias, while 'I', a statement-prone bias, and 'he', neutral. The results showed that this pragmatic bias contributed significantly to the perception of the intonation type. An encouraging finding, however, comes from Ladd *et al.* (1985); they found significant effects of factors of speaker and text, but virtually no interactions between these factors, and acoustic variables studied (i.e., upward/downward trend of f0 contour and f0 range) in their study. So we may generalise findings on the effects of acoustic properties on speaker variables regardless of voices and text used. However, the content effects could more strongly influence ratings of speaker variables, judgements on which tend to be much less clear-cut than such categorical judgements as those of the intonation type. Therefore selection of test passages used for stimuli must be carefully considered so that the utterances do not introduce any noticeable bias.

### **3.5.2.3. Context effects**

In addition to the factors mentioned above, context seems to affect subjects' responses as well. In order to investigate the effects of context on ratings of speaker variables, Brown, Giles and Thakerar (1985) conducted an interesting experiment. They gave half of the subjects no context, while they gave the other half the following context. The context was that the recording was a clip from a recorded lecture by a psychologist to a group of dental students on communication between dentists and children. Subjects were then asked to make judgements on stimuli whose speech rate was manipulated by a human speaker, on 15 scales of speaker variables. They found significant context effects on the intelligence and ambition axes; the subjects who were given the context did not rate slow utterances less intelligent nor less ambitious any longer. So the context effect on rating tasks could be a substantial one.

In the present study, the speaker variable studied is politeness, and politeness judgements are generally very difficult to separated from situations. Therefore, all listener-judges taking part in the perception experiments described in Chapter 6 were informed of the same situations as those which were given to the speakers of the stimulus utterances in the recording sessions, instead of being asked to make judgements on general politeness.

#### **3.5.2.4. Listener-judges**

Although the attributes of listener-judges has received less attention compared with other factors such as listening stimuli and rating methods, this factor is also very important and the influence could be substantial. Judgements are made based on the listener-judge's own evaluation systems, which vary depending on various factors including the listener-judge's sex, age and accent. For example, the sex difference in terms of skill of handling nonverbal cues has attracted a great deal of interest and has been investigated. It appears to be a well established fact that women are superior to men in both decoding and encoding nonverbal cues (e.g., Hall, 1978). Different accents have also been found to affect ratings: different accent groups showed different preference in terms of speaking style in politeness judgement (Section 6.1.1).

The listener attributes mentioned above are social or linguistic ones, but there are other types of factors (e.g., individual-specific factors) which could also influence people's judgement. The potentially relevant factors include the listener-judge's expectations and motivation to complete the task, special skills, training and experience, emotional state, attitude and personality (e.g., Ekman *et al.*, 1980; Rosenthal, 1982, pp. 299-300). Ideally, all the potentially relevant factors should be controlled in perceptual experiments, but practically it is almost impossible. The physical and social factors (e.g., sex, age and accent) are relatively easy to control by



selecting certain groups, and the motivation factors could be controlled reasonably well by, for example, rewarding subjects with payment or credits if they are students, but it is generally very difficult to assess and control the psychological state of the subjects. However, since the effects of these psychological factors are expected to be minor in politeness judgements, compared with those of sex, age and accent, these psychological factors were not controlled in the experiments.

Apart from these social, dialectal and psychological factors mentioned above, there is another factor which was found to be important in interpreting ratings: the acoustic characteristics of the listener-judges' own speech. In fact, a significant correlation between the speech rate of listener-judges and their rate preference was found, and this will be discussed in Section 6.2.

#### **3.5.2.5. Statistical considerations**

In the previous sections we have examined several factors which could influence people's judgements, and seen that ratings do vary depending on these factors. However, many components of individual evaluation systems are expected to be shared by members of the same social or linguistic groups (e.g., sex, age, accent, educational and social backgrounds) because, otherwise, speech communication cannot take place effectively, and this shared knowledge is the main focus of studies on speaker variables. So, although a high degree of agreement between judges' ratings is generally expected, assessing the agreement or reliability of rating scores explicitly is always very useful. It is especially so when low agreement scores were obtained: they may show the inadequacy of the design of rating tasks, or sometimes new hidden factors which should be taken into consideration for interpreting the results.



The reliability of judges' scores can be assessed by correlation coefficients between scores rated by different judges. The correlation coefficients commonly used are the product-moment correlation coefficients (e.g., the Pearson  $r$  and Spearman's rank-order correlation coefficient), Kendall's tau coefficient and Kendall's coefficient of concordance ( $W$ ) (e.g., Howell, 1992, Chapter 10). All the coefficients mentioned above except Kendall's  $W$  deal with the relationship between two sets of scores, so when more than two judges are involved, the mean value of correlation coefficients between all possible pairs of rankings is usually used as an index of the reliability of a single average judge (e.g., Rosenthal, 1982, pp. 292-299). Kendall's  $W$ , which is defined as the ratio of the total variability among different judges' rankings for each stimulus to the maximum possible variability, is a measure of degree of association between more than two sets of ranking scores, and bears a linear relation to the mean Spearman's rank-order correlation coefficient (Howell, 1992, pp. 280-282).

These measures mentioned so far concern the reliability of a single average judge. Rosenthal (1982, pp. 292-299) suggests that the reliability of the ratings of all the judges involved (which is called 'effective reliability' as opposed to 'mean reliability') should also be reported for assessing the reliability of using listener-judges' ratings as the definition of the encoder's state or nonverbal behaviour. The effective reliability ( $R$ ) can be estimated in various ways. If the mean reliability ( $r$ ) is already available,  $R$  can be calculated by employing the Spearman-Brown formula, (i.e.,  $rn / [1 + (n-1)r]$ , where  $n$  is the number of judges) (Rosenthal, 1982, p. 293). Another way to estimate  $R$  is to use the analysis of variance (ANOVA) on the ratings scores themselves. This method calculates two different estimates of the population variance: one is based on the assumption that the mean values for the ratings of each category (e.g., stimuli, encoders) are the same (MS category), and the other is independent of this assumption (MS error).  $R$  can be directly estimated by employing the following formula:

$$[MS(\text{category}) - MS(\text{error})] / MS(\text{category}) \quad (\text{Rosenthal, 1982, p. 296}).$$

Inter-judge agreement may vary depending on various factors concerning the listener-judges, but if intra-judge agreement is high, using the rating scale method is still justifiable. It can be assessed by correlation coefficients between test-retest scores of the same listener-judges. Using the test-retest method, however, needs caution because of a learning effect on the second judgements, and may not be practical because of, for example, unavailability of the same person. Another way to assess the reliability of each judge's ratings is to calculate the ratio between variance of each judge's repetition scores (i.e., rating scores of the same stimulus rated by the same judge), and the total variance, by performing the ANOVA. If this repetition factor is found to be non-significant, it can be concluded that each judge's ratings are reasonably consistent.

In the present study, the inter-judge agreement was assessed by Kendall's W (hence the mean Spearman rank-order correlation coefficient), and the intra-judge agreement was assessed by the significance of the judges' repetition factor. Since the effective reliability ( $R$ ) is very important to assess using listener-judges scores for stimulus evaluation in terms of the effectiveness of cue manipulation,  $R$  will also be reported in the perceptual experiments (Chapter 6). As we have seen, the underlying assumption of the concept of the effective reliability is that the differences between individuals do not have significant importance in their ratings, and hence can be cancelled out if a reasonably large number of judges' scores are used. However, such individual differences as linguistic and social backgrounds do usually have significant effects on the ratings, and therefore could be a very important factor to understand human perception of speaker variables. So both the effectiveness of cue manipulation assessed by all judges' scores and differences between judges will be discussed in Chapter 6.



### 3.6. Summary

This chapter discussed various issues which should be taken into consideration for conducting research on vocal features in relation to speaker variables. Approaches used in studies regarding speaker variables were reviewed, and among them, the combination of acoustic analysis and perceptual experiments with computer cue manipulated stimuli was judged to be the best possible approach for the present study. The findings of studies which focused on paralinguistic features in relation to Japanese politeness showed that three factors, (i.e., pitch, tempo and loudness), could be potential cues, and acoustic variables related to these perceptual variables and voice quality investigated in relation to speaker variables were reviewed, in order to determine what aspects of vocal features to measure and manipulate in searching for acoustic cues to politeness. Then three factors were discussed with regard to speech data collection: selection of elicitation methods, selection of test passages, and speaker variability. Among these considerations, selection of elicitation methods is known to be very important, because it involves an essential question: whether or not laboratory recordings are natural enough for studying speaker variables. Although naturalness is a vital factor, due to the difficulty of obtaining a high degree of control on situational factors in field recordings, a compromise approach, which attempts to achieve natural simulations under laboratory conditions, was introduced as practically the best method. Next, methodological issues concerning computer cue manipulation techniques were discussed. Despite ecological validity problems, which concern whether or not computer manipulated utterances are realistic enough, the great advantages of using computer cue manipulation over human manipulation were argued. Several computer cue manipulation techniques were reviewed and the PSOLA technique was found to be the best for the stimulus preparation in this study. Finally, several factors which could influence people's ratings, and statistical considerations regarding rating tasks were discussed.



### **3.7. Approach adopted in the present study**

Based on the reviews in this chapter, the following approach was adopted to investigate the acoustic cues to Japanese politeness. This approach consists of three stages:

- (1) record stimulus utterances which convey different degrees of politeness (Chapter 4);
- (2) conduct a preliminary investigation of acoustic variables which appear to be relevant to the degree of politeness conveyed in the recorded utterances (Chapter 5); and
- (3) conduct perceptual experiments with stimuli which were created by computer cue manipulation of potentially relevant acoustic variables, using selected utterances from Stage (1) as source utterances, in order to observe the effects of these manipulated variables on politeness judgements (Chapter 6).

## CHAPTER 4

### SPEECH DATA COLLECTION

This chapter describes the elicitation method used for collecting polite and non-polite utterances. The utterances were then evaluated by a panel of listener-judges in order to examine which utterances were good representatives of politeness at different levels.

#### 4.1. Politeness elicitation

As we have seen in the previous chapter (Section 3.4), there is always a trade-off between the realism of field recordings and a high degree of control in laboratory recordings. A role-play method in which speakers are given scenarios and asked to play their roles (e.g., Hong, 1993) was adopted as a compromise, in order to obtain natural simulations under laboratory conditions.

Sentences with 'semantically neutral' content were used as test passages because politeness judgements cannot be separated from situations, and therefore the content would be an indispensable part of the judgement. The sentences used were:

- (i) *Nimotsu-wa koredake desuka*, meaning 'is this all the luggage you have?'; and
- (ii) *Moshimoshi Akagi-san<sup>1</sup> no otaku desuka*, meaning 'hello is that Mr. Akagi speaking?'. The 'luggage' sentence is a routine question usually heard at the customs office, and the 'hello' sentence is a conventional expression at the beginning of telephone conversations. The sentences were selected because they are so commonly used and so conventional that listeners should pay minimal attention to the content.

---

<sup>1</sup>: the suffix '-san', although translated into English as 'Mr', does not have the same politeness connotations. It has a much more general usage, contrasting with more specific forms of address, e.g., 'Professor', and with address to animals and young children.

Since the concept of politeness is closely related to appropriateness in a situation, there can be various realisations of both politeness and impoliteness. The same utterance can be judged as both polite and non-polite depending on the situation. So it is very important to determine which aspects of politeness or impoliteness are to be investigated. As we have seen in Chapter 2, Japanese politeness is almost always associated with the honorifics, which are used for the speaker's seniors or superiors, and the honorific system indeed functions as a relation-acknowledging device in Japanese society (Matsumoto, 1988). Since this relation-acknowledging function is so important in any kind of social interaction, people know how to speak appropriately in a given situation. However, they appear to have difficulties in speaking prosodically politely or impolitely without any situational context unless they are very experienced speakers, and using experienced speakers is not always desirable because of their highly theatrical or stereotyped expressions. Therefore, scenarios were used to induce different levels of politeness by giving the speakers situations.

The categorisation of 'situation' by Brown and Fraser (1979) is shown in Fig. 4.1. They categorise situational factors into two groups: scene and participants. The scene factors consist of 'setting' and 'purpose'. The participant factors consist of 'individual participants' (e.g., personality and emotions) and 'relationships between participants', which is further divided into two factors: 'interpersonal relations' (e.g., liking, knowledge) and 'role and category relations' (e.g., social power and status). Among these situational factors, the role relationship is the most influential factor for determining politeness levels in Japanese because of close ties between politeness levels and the category of addressee in Japanese, as was seen in Chapter 3 (e.g., Hill et al, 1986).



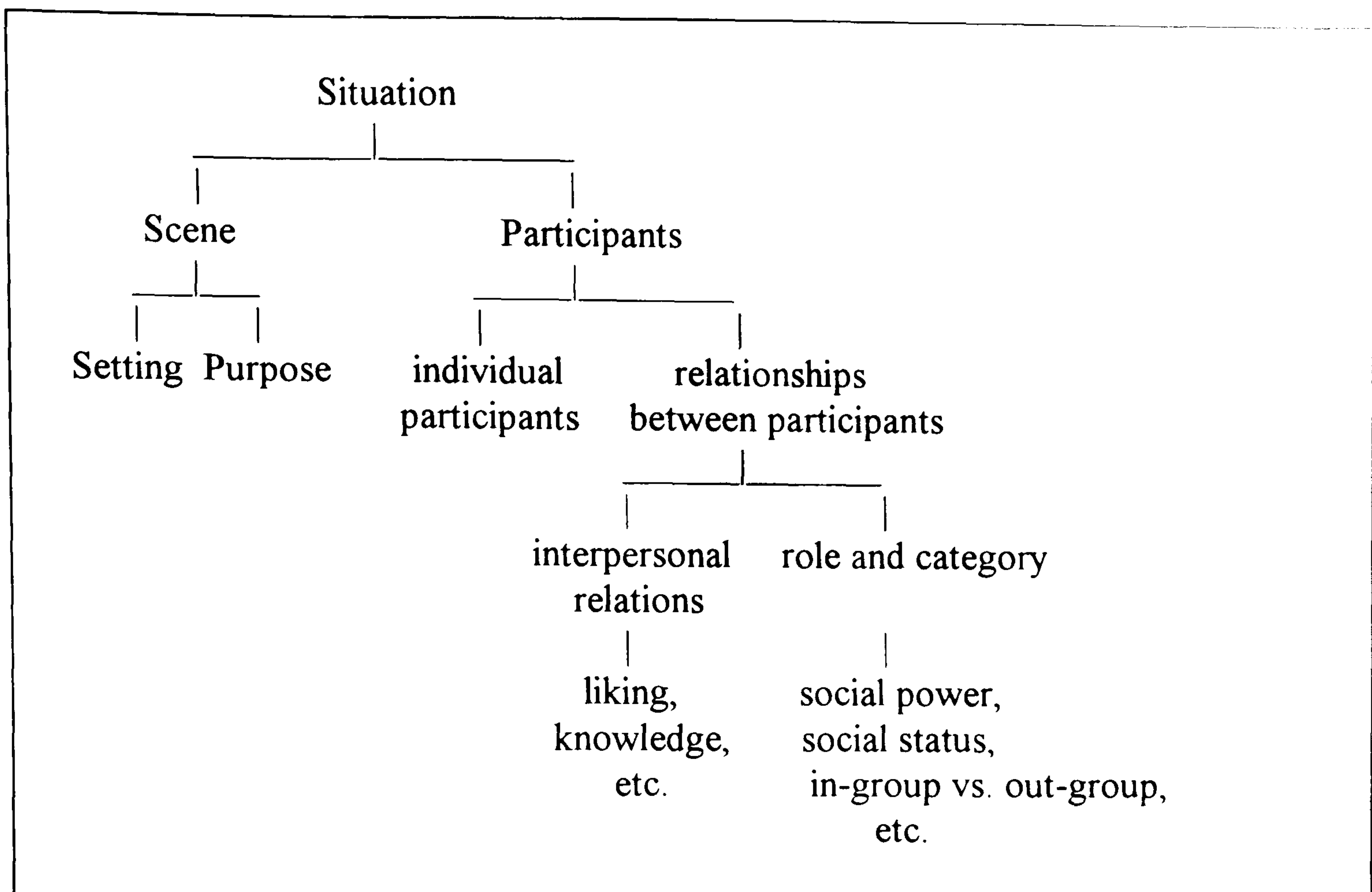


FIG. 4.1. Components of situation. (A simplified figure of the original figure in Brown and Fraser (1979, p. 34)).

The scenarios used were: (1) A young customs officer talking to three types of passengers (a respectable gentleman, a casually dressed young student, or a shabby drunk) at an airport; and (2) A young public officer talking to the same three types of citizens on the telephone. Utterances addressed to a respectable gentleman are expected to be polite, whereas utterances addressed to either a young student or a drunk were expected to be non-polite: casual or uninhibited to a student, and authoritative to a drunk. Although there is a strong association between the type of addressee and speaking styles in Japanese speech, the type of addressee does not necessarily induce the same speaking style from speakers. So the desired speaking styles (i.e., 'polite', 'casual' and 'authoritative') were specified in the instructions (see Appendix A). The test passages were located at the end of each short conversation between the speaker and addressee.

A total of six untrained native speakers of Japanese in their twenties, who could speak standard Japanese, participated in the recording sessions of about two hours. Four of the speakers were undergraduates at a Japanese college in England and the other two were postgraduate students at a British university. The age and hometown of the speakers are shown in Table 4.1. All the speakers were paid for their participation.

TABLE 4.1. Age and hometown of the speakers.

<i>Speaker</i>	<i>Age</i>	<i>Hometown</i>
TN	22	Tokyo
KS	22	Fukuoka
HA	22	Osaka
TK	25	Yokohama
SF	26	Niigata
KI	29	Niigata

The speakers were given descriptions of the two situations and the characteristics of the speaker (a customs officer in Scenario 1 and a public officer in Scenario 2) and the three types of addressee. In order to help speakers get into the roles, model dialogues of short conversations between a speaker and addressee were also provided. The instructions, the scenarios and the model dialogues given to the speakers, and their English translations are attached in Appendix A. The speakers, who knew each other well, worked in pairs, one playing the role of speaker and the other the role of addressee. After having a practice session of about one hour, the short conversations and the test passages (which the speakers were requested to speak two or three times) were recorded. After the recording in each situation characterised by the type of addressee, subjects exchanged the roles and repeated the same procedure. At the end

of the session, subjects were asked to talk to their partner about themselves (e.g., what they were studying) and their impressions of the role-play in order to measure the normal pitch range of their voice. The recordings were made on digital audio tape with a SONY TCD-D7 DAT recorder and a SONY F-V07T dynamic microphone, and then digitised and stored at a sampling frequency of 16 kHz onto a Sun workstation via its own A-to-D boards.

## 4.2. Utterance evaluation

The utterances which were obtained in the recording sessions described in the previous section were first evaluated by informal auditory assessment by the author. It was confirmed that each utterance which was meant to be polite sounded more polite than utterances which were intended to be casual or authoritative, spoken by the same speaker, whereas the differences between casual utterances and authoritative utterances were not very clear in many cases. Most of the speakers did not use an authoritative tone to a 'drunk', but used either a casual or soothing tone. The utterances were then evaluated by a panel of listeners in terms of how well each utterance achieved the intended politeness levels (i.e., politeness, casualness and authoritativeness).

### 4.2.1. Method

#### *1. Listening material*

The listening material is the utterances of the two sentences (i.e., the 'luggage/*nimotsu*' sentence and the 'hello/*moshi*' sentence) spoken by the six speakers with the three speaking styles: 'polite', 'casual' and 'authoritative'. Among several utterances spoken by the same speaker for the same style and sentence, the first



utterances were used because they sounded most natural; speakers tended to 'recite' their subsequent utterances.

## *2. Stimulus presentation and evaluation procedure*

The material is divided into six sets, each of which consists of utterances of the same sentence and the same speaking style, spoken by the six speakers. A paired comparison method for the six conditions of each set (i.e., six speakers) was used to assess how well each utterance represents the intended speaking style (i.e., politeness, casualness or authoritativeness) (Watkins and Makin, 1994). Before the session started, subjects were given written instructions, which included the situations given to the speakers (see Appendix B). After a short practice session, scores for politeness, casualness and authoritativeness were collected in this order. In the 'politeness' session, first, subjects were presented with 'polite' utterances of the 'luggage' sentence, and after a short break, the sessions with 'polite' utterances of the 'hello' sentence followed. On each trial, subjects heard two utterances successively preceded by a warning tone, and selected which utterance sounded more polite to them. In each sentence set (i.e., the 'luggage' sentence or 'hello' sentence) each utterance was compared with every other utterance, using both orders of presentation; hence it yields 30 paired comparisons. Subjects were presented with a total of 32 trials including two dummy trials at the beginning and the end of each session, randomised differently for each subject, through headphones. The sessions for casualness and authoritativeness were run the same way. The whole session took about 30 minutes.

## *3. Subjects*

There were five paid subjects (2 male and 3 female). All subjects were native speakers of Japanese, ranging in age between 21 to 27, and were postgraduates at a British university.

4.2.2. Results and discussion

The scores were calculated as the number of times an utterance was judged more polite/casual/authoritative in a comparison, divided by the total number of occurrences of each utterance. The scores could range from 0 (least polite/casual/authoritative) to 1 (most polite/casual/authoritative), that is they are ordinal data.

Inter-judge agreement was assessed by Kendall's coefficient of concordance (W) for each set (Table 4.2). The results showed a reasonably high agreement between the listener-judges' scores. A slightly higher level of agreement was found for the 'hello' sentence.

TABLE 4.2. Kendall's coefficient of concordance (W) between five listener-judges' scores for six different speaker conditions. The significance is at the level of 0.05 or better.

<i>Speaking Style</i>	<i>Sentence</i>	
	<i>LUGGAGE</i>	<i>HELLO</i>
Polite	0.62	0.71
Casual	0.59	0.63
Authoritative	0.67	0.66

The mean values for the scores of utterances of the two sentences spoken by the six speakers, with the three speaking styles are shown in Table 4.3. The scores are arranged from the best to the least representative of the style. The results show that no single speaker was judged as the best encoder for all the speaking styles: SF was the best for politeness, KS for casualness and TK and HA for authoritativeness. However,

the speakers' performance is fairly consistent across different sentences (i.e., good encoders for one sentence are also good for the other). Among these six speakers, TK and KS were generally judged as good encoders for these three speaking styles. The waveforms and f0 contours of the 'polite' and 'casual' utterances of both the 'luggage' and 'hello' sentences spoken by three speakers (KS, TK and SF) are attached in Appendix C.



TABLE 4.3. Mean scores across five listener-judges' scores in the utterance evaluation test. The scores for politeness (a), casualness (b) and authoritativeness (c) are shown separately. 'Speakers' are ordered from the highest score to the lowest.

(a) POLITENESS

<i>'Luggage' sentence</i>		<i>'Hello' sentence</i>	
<i>Speaker</i>	<i>Mean</i>	<i>Speaker</i>	<i>Mean</i>
SF	0.86	SF	0.90
TK	0.66	KS	0.74
KI	0.48	TK	0.48
HA	0.44	HA	0.32
TN	0.36	TN	0.32
KS	0.20	KI	0.24

(b) CASUALNESS

<i>'Luggage' sentence</i>		<i>'Hello' sentence</i>	
<i>Speaker</i>	<i>Mean</i>	<i>Speaker</i>	<i>Mean</i>
KS	0.88	KS	0.92
TN	0.66	TK	0.56
TK	0.46	TN	0.44
HA	0.44	SF	0.44
SF	0.28	HA	0.40
KI	0.28	KI	0.22

(c) AUTHORITATIVENESS

<i>'Luggage' sentence</i>		<i>'Hello' sentence</i>	
<i>Speaker</i>	<i>Mean</i>	<i>Speaker</i>	<i>Mean</i>
TK	0.84	TK	0.90
HA	0.78	HA	0.76
KS	0.54	SF	0.48
KI	0.30	KS	0.38
TN	0.28	KI	0.32
SF	0.26	TN	0.16

## CHAPTER 5

### ACOUSTIC ANALYSIS

This chapter describes acoustic analysis of polite and non-polite utterances spoken by the six male Japanese speakers. The details of the material are described in the previous chapter. The purpose of this acoustic analysis was to investigate distinct acoustic variables which could be identified as consistently distinguishing speaking styles with different politeness levels.

#### 5.1. Acoustic features chosen for measurement

F0 and temporal variables form the focus of this study for several reasons: first, they are major acoustic correlates of perceived pitch and tempo, which are considered to be important by Japanese people in politeness judgement in Japanese speech (Ogino and Hong, 1992); second, they are robust in the sense that they survive even in very noisy environments and through degraded telephone lines; third, they are relatively easy to measure and manipulate by means of currently available computer software; and finally, appropriate settings of both duration and f0 were found to change the perception of politeness (Ofuka *et al.*, 1994).

The two sentences (the 'luggage' and 'hello' sentences) spoken by the six male Japanese speakers in the three speaking styles ('polite', 'casual' and 'authoritative'), which are described in Chapter 4, were used in the acoustic analysis. The most natural utterance among utterances of the same sentence spoken by the same speaker in the same speaking style recorded on digital audio tape was selected for each speaker, and digitised at a sampling rate of 16 kHz onto a Sun workstation via its own A-to-D boards.

The two sentences used in this acoustic analysis can be divided into two parts: the 'luggage' sentence consists of 'nimotsu-wa' (Phrase 1) and 'koredake desuka' (Phrase 2); the 'hello' sentence consists of 'moshimoshi' (Phrase 1) and 'Akagi-san no otaku desuka' (Phrase 2). A pause could be inserted between Phrase 1 and Phrase 2. F0 values and segmental durations of utterances of these sentences were measured using the digital processing software package ESPS/Waves (Entropic Research Laboratory, 1993). Acoustic variables measured were mean f0 (in Hz), coefficient of variation (SD divided by mean), range (the difference between the 95th percentile point and the 5th percentile point in semitones), rate of change in f0 (the slope of fitted regression lines for f0 contours in each mora), duration of total utterances and pauses (in milliseconds or ms), articulation rate exclusive of pauses and final mora duration (in mora per second), the duration of the final morae (in ms) and f0 final directions (rise or fall) and steepness (in semitones per second, or st/sec). The unit 'semitone' was adopted for f0 range in order to compare ranges of different speakers with different ranges of voice. A distance (D) in semitones between two frequencies measured in Hz (f1 and f2) is calculated with the following formula:  $D \text{ (in semitones)} = [12 / \text{LOG } 2] \times \text{LOG } (f1/f2)$  (where LOG is a logarithmic function with a base of 10). 'Mora' corresponds to a Japanese phonetic syllable, basically consisting of either a vowel (V) or a consonant followed by a vowel (CV) with two exceptions (i.e., a syllabic nasal /N/ and a unit of silence called '*sokuon*'). The mora is said to be equal to "the full length of a short syllable or half the length of a long syllable" (Cruttenden, 1986, pp. 13-14). The durations of final morae were excluded for calculation of articulation rate because of the great variability found among the samples.

Among the f0 contour variables discussed in Section 3.3.3.1, final f0 movement was particularly focused on in the present study because of the recognition of the importance of its role as an information carrier in the literature. The final part is also



very important for expressing affect in terms of linguistic form in Japanese, as final particles ('shuujoshi'), which signal attitudinal meanings of the speaker, are located at the end of the sentence. Furthermore, the final part has greater freedom in  $f_0$  movement compared with the other part of the sentence. In Japanese, pitch accent is fundamental to defining word identity in much the same way as it is in tone languages; hence speakers of Japanese have limited freedom in terms of the shape of  $f_0$  contours except the final part.

Although voice quality and articulation were not the main focus of the present study, auditory assessment and spectral analyses using wide and narrow bandwidth spectrograms were performed. These were to investigate (1) whether or not there are noticeable differences in voice quality and articulation, with different politeness levels, and (2) if there are, what acoustic variables could be relevant to these differences.

## **5.2. The outcome of acoustic analyses**

The mean values for the  $f_0$  and temporal variables for the polite and casual versions (including the final morae) of the two sentences are shown in Tables 5.1 and 5.2. Table 5.3 shows the duration and  $f_0$  rate of change of the final morae alone in Phrase 1 and Phrase 2. The comparisons between these acoustic variables in the polite versions and those in the casual versions are also shown in Tables 5.4 and 5.5. Since auditory assessment showed no clear difference between the casual and authoritative versions for most of the speakers, the authoritative versions were excluded from these summary tables. The measurements of these acoustic variables in each speaker's utterances of the two sentences and natural conversations which were recorded at the end of the recording sessions are attached in Appendix D. The measurements of the authoritative versions are included only for speakers TK and HA, who were judged as the best encoders by a panel of listeners (Section 4.2).

The  $f_0$  related variables selected for this acoustic analysis were mean  $f_0$ , range and steepness. The range was assessed by the difference between the 95th percentile point and the 5th percentile point in semitones. The steepness was assessed by the mean values for absolute values of regression coefficients fitted to the  $f_0$  contours. Before the calculation of these indices, extreme values (e. g., octave jumps) were manually eliminated. Table 5.1 shows that these  $f_0$  variables (i.e., mean  $f_0$ , range and steepness of  $f_0$  contours) were not significantly different in the polite and casual speaking styles (in fact, the values were almost equal), and great variability is apparent in usage of these  $f_0$  variables among the six speakers (Table 5.4); hence the importance of the  $f_0$  variables studied as a cue for signalling politeness is inconclusive. On the other hand, the temporal variables (i.e., articulation rate, utterance length and the final vowel duration) showed significant difference in different styles at least in one sentence (Tables 5.2 and 5.3) and are also consistent across the six speakers (Tables 5.4 and 5.5). Measurements of each acoustic variable, and auditory and spectral analyses for potential differences in voice quality and articulation are discussed in the following subsections.

TABLE 5.1. F0 variables in polite (P) and casual (C) utterances: mean values and SDs across six male speakers.

<i>ST</i>	<i>Mean f0</i> <i>(in Hz)</i>		<i>Range</i> <i>(95% - 5%)</i> <i>(in semitones)</i>		<i>Steepness*1</i> <i>(mean value for regression coefficients)</i>	
	P	C	P	C	P	C
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
1	135.5 (13.7)	140.4 (17.4)	12.89 (2.83)	12.14 (4.45)	3.25 (1.50)	3.30 (0.88)
2	152.6 (17.2)	151.5 (18.1)	10.71 (2.63)	11.73 (4.46)	2.32 (0.68)	2.15 (0.93)

ST 1: 'luggage' sentence and ST 2: 'hello' sentence.

\*1: regression coefficients were calculated with normalised f0 values of each speaker. All f0 values of each speaker were normalised in such a way that the lowest f0 and the highest f0 of the speaker is 0 and 100. The lowest and highest f0 were among f0 values of all the utterances of the two sentences and the natural conversations of the speaker. Steepness was assessed by mean values for absolute values of regression coefficients.

NB: all the difference between mean values for the polite and casual styles are non-significant by 2-tailed, paired t-test at  $p = 0.05$ .



TABLE 5.2. Temporal variables in polite (P) and casual (C) utterances: mean values and SDs across six male speakers.

<i>ST</i>	<i>Speech rate*1</i> <i>(in mora/sec)</i>		<i>Total utterance</i> <i>length</i> <i>(in ms)</i>	
	P	C	P	C
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
1	11.37 (0.61)	< 12.90 (1.14)	1113 (92)	> 1040 (125)
2	12.65 (0.82)	< 13.40 (0.48)	1760 (292)	1695 (319)

\*1: the duration of a pause between Phrase 1 and Phrase 2, and the final morae in Phrase 1 and Phrase 2 were excluded from the calculation of speech rate.

ST 1: 'luggage' sentence and ST 2: 'hello' sentence.

< and > show the relationship between the mean value of the polite versions (P) and that of the casual versions (C), and the difference is significant by 2-tailed, paired t-test at  $p = 0.05$ .

TABLE 5.3. Final morae: duration and f0 rate of change in Phrase 1 and Phrase 2 of poilte(P) and casual (C) utterances. Values represent mean values and SDs across six male speakers.

ST	Mora: wa/shi (in Phrase 1)				Final vowel: a (in Phrase 2)	
	Duration (in ms)		f0 rate of change (in semitones/sec)		Duration (in ms)	
	P	C	P	C	P	C
	Mean	Mean	Mean	Mean	Mean	Mean
	(SD)	(SD)	(SD)	(SD)	(SD)	(SD)
1	100.0 (47.7)	65.0 (12.2)	66.0 (35.2)	<* 113.8 (25.9)	91.7 (33.1)	138.3 (61.8)
2	193.3 (43.2)	>? 163.3 (52.8)	54.2 (25.0)	53.6 (23.0)	100.0 (40.5)	<* 163.3 (53.5)

ST 1: 'luggage' sentence and ST 2: 'hello' sentence.

< and > show the relationship between the mean value of the polite versions (P) and that of the casual versions (C). '\*' means that the difference is significant by 2-tailed, paired t-test at  $p = 0.05$ , while '?' means that the difference just failed to reach significance ( $p = 0.07$ ). The other differences were not significant at  $p = 0.05$ .

TABLE 5.4. Comparisons between f0 and temporal aspects of polite versions (P) and those of casual versions (C) of two sentences spoken by six male speakers.

		<i>F0</i>				<i>DURATION</i>	
		<i>mean</i> <i>(Hz)</i>	<i>SD/mean</i>	<i>range</i> <i>95% - 5%</i> <i>(semitones)</i>	<i>rate of</i> <i>change</i> <i>*1</i>	<i>speech rate</i> <i>(mora/sec)</i>	<i>total</i> <i>length</i> <i>(sec)</i>
<i>Speaker</i>	<i>ST</i>	P C	P C	P C	P C	P C	P C
TN	1	<	<	<	<	<	>?
	2	<?	<	<	<	<	<?
KS	1	<	<?	>	<	<	>?
	2	<?	<	<	<?	<	>
HA	1	>	>	>	>	<	>
	2	>	<	<	<	<?	<?
TK	1	<	>	>	<	<	>
	2	>	>?	>	>	<?	<?
SF	1	<?	<?	>	<	<?	>
	2	<?	>	>	>	<?	>
KI	1	<?	<	<	>	<	>
	2	<	<	<	<	<?	>?

\*1: rate of change was assessed by mean values for absolute values of regression coefficients.

ST 1: 'luggage' sentence and ST 2: 'hello' sentence.

<, > and ? (the difference is less than 5% of the minimum value of the two) show the relationship between the value for the polite version (P) and the value for the casual version (C) spoken by the same speaker.



TABLE 5.5. Comparisons between the characteristics of the final mora of the first and the second phrase of the sentence of polite versions (P) and those of casual versions (C) of two sentences spoken by six male speakers.

		<i>Mora: wa/shi</i> <i>(in Phrase 1)</i>		<i>Final vowel: a</i> <i>(in Phrase 2)</i>	
		<i>duration</i> <i>(ms)</i>	<i>f0 rate of</i> <i>change</i> <i>(semitones/s)</i>	<i>duration</i> <i>(ms)</i>	<i>f0 direction</i> <i>(f0 rate of change in</i> <i>semitones/sec)</i>
<i>Speaker</i>	<i>ST</i>	P C	P C	P C	P C
TN	1	>	<	<	/ (24)      _ / (4; 114)
	2	=	>	=	∨ (-147; 31)    \ (-34)
KS	1	=	<	<	\ (-54)      _ / (-9; 56)
	2	>	>	<	\ (-41)      _ (-7)
HA	1	<	<	<	f0 tracking error
	2	=	>	<	\ (-58)      ∨ (-38; 32)
TK	1	>	<	<	/ (28)      / (20)
	2	=	devoiced	<	/ (29)      _ (3)
SF	1	>	<	<	/ (55)      / (26)
	2	>	>	<	/ (21)      / (60)
KI	1	>	<	>	/ (38)      / (21)
	2	>	<	<	_ (-6)      / (10)

ST 1: 'luggage' sentence and ST 2: 'hello' sentence.

<, > and = show the relationship between the value for the polite version (P) and the value for the casual version (C) spoken by the same speaker.

'/' means a rising tone (values  $\geq 10$  semitones/sec), '\', a falling tone (values  $\leq -10$  semitones/sec), and '\_', a level tone ( $|\text{values}| < 10$  semitones/sec). '∨(n1, n2)' and '\_/(n1, n2)' mean that the f0 movement consisted of two lines, and the steepness of these lines are in parentheses (i.e., n1 for the first line and n2 for the second line).

### 5.2.1. F0 level

The level of  $f_0$  was assessed by calculating the mean values of  $f_0$ . Although the level of  $f_0$  has received a great deal of interest in relation to politeness in Japanese speech as we have seen in Section 3.3.2, whether the  $f_0$  level can be a cue for politeness remains inconclusive especially for male speakers. The results of the measurement show great variability among speakers in use of  $f_0$  levels for the polite and casual styles (Table 5.4); hence this does not seem to be an important factor for distinguishing these styles. Speakers TN and KS adopted higher voice for casualness while SF and KI did not change the level of voice, for both sentences. Speaker TK, who was judged to be a good encoder among the speakers in the utterance evaluation test described in Section 4.2, adopted higher pitch for casualness for the 'luggage' sentence, but higher pitch for politeness for the 'hello' sentence. The absolute level of  $f_0$  does not seem to be a cue either: Speaker SF, whose polite utterances were judged to be the most polite, used a  $f_0$  level which is somewhere in the middle among the six speakers (Table D.1 in Appendix D).

However, the great variability found among the six speakers in their use of  $f_0$  levels in different styles does not necessarily mean that the level of  $f_0$  cannot be a sign of politeness, because the variability here is the variability found in the comparison between the use of  $f_0$  levels for politeness and that for casualness. High-pitched voice can also be used to express familiarity as well as politeness. In fact, Ogino and Hong's (1992) survey shows that fast, high-pitched or 'normal', and strong speech was for close friends. So perhaps some of the speakers who took part in our recording sessions adopted high-pitched voice for expressing familiarity in the 'casual' situation, in which they were supposed to speak to a young student.

Since Ogino and Hong's (1992) survey also showed that high-pitched voice was considered to be used for their superiors and strangers by some of the informants, as well as for close friends, Ogino and Hong conclude that high-pitched voice probably shows that the voice is 'marked', (i.e., the speaker is aware that the situation is not ordinary). Speech for superiors and strangers can be 'marked' because of formality or tension, and speech for close friends can also be 'marked' because of excitement.

In order to investigate this claim (i.e., the association of higher pitch and markedness), the mean  $f_0$  of utterances of the two sentences spoken in both polite and casual ways was compared with the mean  $f_0$  of each speaker's utterances in a one or two minute dialogue with his recording partner, which was recorded at the end of the session in a relatively relaxed atmosphere. It was found that the mean values over the six speakers for the 'polite'/'casual' utterances (about 140 Hz for the 'luggage' sentence and about 150 Hz for the 'hello' sentence) were significantly higher than the mean value over the speakers for the natural utterances (120 Hz) (2-tailed t-test,  $ps < 0.05$ ) (Table 5.1). The speakers did adopt relatively higher voice for 'polite' and 'casual' utterances compared with the ordinary voice. Although it is not entirely clear that high-pitched voice was used especially for politeness and casualness, or this merely showed that the speakers were using the 'acting' voice, or only stressed in the recording sessions, this result supports such hypothesis that high-pitched voice is a sign of markedness. It was also found that the mean  $f_0$  for the 'hello' sentence was significantly higher than that for the 'luggage' sentence for both speaking styles (2-tailed t-test,  $ps < 0.01$ ), perhaps showing that there is the 'telephone' voice.

In summary, although the level of  $f_0$  alone does not appear to be an absolute cue because of the inconsistency in usage in the polite and casual styles, it can be concluded that higher pitch is associated with markedness, and therefore, could be one



of the signs of politeness (e.g., by signalling that the speaker is tensed because of the presence of an important addressee).

### **5.2.2. F0 variability, range and rate of change**

F0 variability was assessed by coefficient of variation (SD divided by mean), and f0 range by the difference between the 95th percentile point and the 5th percentile point measured in semitones per sec. The rate of change in f0 was assessed by calculating the mean value for the absolute values of slopes of the regression lines (regression coefficient) fitted to the f0 contour (smoothed by hand) of each mora. In the calculation of the regression coefficients, normalised f0 values were used: all f0 values of each speaker were normalised in such a way that the lowest and highest f0 among the f0 values of all the utterances of two sentences and the natural conversations of the speaker are 0 and 100. Table 5.4 shows that use of all these variables for the polite and casual styles varied from speaker to speaker, and even from sentence to sentence depending on the speaker. Therefore it is very unlikely that these acoustic variables are used intentionally by speakers as a cue for politeness.

However, there was one noticeable difference in the use of f0 variability/range found in Speaker TK's authoritative version. TK, whose authoritative version was judged as the best representative among the six speakers, adopted an extremely narrow range to express authoritativeness: the range was less than 30% of the ranges which were adopted for the polite and casual versions (for details, see Table D.1 in Appendix D). Since this extremely narrow range for authoritativeness was not adopted by HA, who was judged as the second best for authoritativeness, using narrow ranges may not be an universal strategy, but this appears to work very effectively together with other appropriate vocal features for authoritativeness.

### 5.2.3. Articulation rate

The tempo, for which articulation rate is the major acoustic variable, has been recognised as the most noticeable factor for politeness by native speakers. According to Ogino and Hong's (1992) survey, the majority of 200 or so informants replied in the interview that slow speech was perceived as polite, and that they would speak slowly to their superiors and strangers. This native speakers' intuition was also supported by the measurement of the articulation rate of polite and casual utterances. Table 5.4 shows that all the speakers consistently adopted slower rate for politeness, and the difference between the mean value for the articulation rate over the six speakers for politeness and that for casualness is significant for both sentences (2-tailed t-tests,  $ps < 0.05$ ) (Table 5.2). Therefore, it can be concluded that the articulation rate can be a very reliable cue for politeness.

### 5.2.4. Total utterance length and pause

The main difference between the total utterance length and articulation rate is the factor of pause. Pauses have been investigated in relation to hesitation, which is often associated with politeness or formality, and thus could be a cue for politeness.

The duration of pauses adopted by each speaker is shown in Table 5.6. This shows that most of the speakers inserted a pause for the polite versions, but not necessarily for the casual and authoritative versions, and the length of pause is longer than that for the casual versions in most cases. So it can be summarised as follows: longer pauses tend to be present in polite speech, but long pauses alone do not necessarily signal high politeness levels.

TABLE 5.6. Pause durations (in ms) in three different speaking styles in two sentences (the 'luggage' sentence (a) and the 'hello' sentence (b)) spoken by six male speakers. The measurements for the authoritative versions are only shown for the best encoders of this speaking style (HA and TK).

(a) 'LUGGAGE' SENTENCE

<i>Speaker</i>	<i>Polite</i>	<i>Casual</i>	<i>(Authoritative)</i>
TN	70	70	
KS	50	20	
HA	0	0	0
TK	10	0	0
SF	110	0	
KI	10	0	

(b) 'HELLO' SENTENCE

<i>Speaker</i>	<i>Polite</i>	<i>Casual</i>	<i>(Authoritative)</i>
TN	520	700	
KS	400	110	
HA	500	490	860
TK	160	150	150
SF	150	0	
KI	0	0	



### 5.2.5. Final f0 movement

The final vowel 'a' of the sentences was analysed in terms of duration, and direction of the f0 movement. The duration of the final vowel was significantly and consistently longer for the casual versions (Tables 5.3 and 5.5). The difference was significant for the 'hello' sentence (2-tailed t-test,  $p = 0.035$ ), but did not reach the significance level of 0.05 for the 'luggage' sentence ( $p = 0.107$ ). It was also found that the ratio of the duration of the final morae ('ka') to the standard duration for each mora (i.e., [total utterance length - pause] divided by the number of morae) was found to be very consistent across the six speakers especially for politeness. For the 'luggage' sentence, the mean value for the ratio was 1.7 (ranging from 1.3 to 2.0) for politeness, and 2.2 (ranging from 1.6 to 2.9) for casualness. For the 'hello' sentence, the mean value was 1.8 (ranging from 1.5 to 2.2) for politeness, and 2.6 (ranging from 2.0 to 3.3) for casualness. The final direction, however, did not show any clear difference depending on the speaking style: all the speakers except KS, adopted a rising tone regardless of the speaking style for the 'luggage' sentence, while no clear pattern was found for the 'hello' sentence. So, in conclusion, the duration of the final vowel can be a cue to politeness while the contribution of the final f0 direction is inconclusive.

The last morae of the first phrase were also measured in terms of duration (in ms) and the steepness of the f0 movement (a fall in all cases) (in st/sec). The duration was found to be longer for politeness, but the difference was insignificant for both sentences (Table 5.3). However, the consonant 'sh' of the last mora 'shi' of the 'hello' sentence, 'sh' was pronounced significantly longer for politeness (2-tailed t-test,  $p = 0.020$ ), which may be related to careful articulation. The differences in articulation are discussed in the next subsection. The steepness was found to be significantly steeper for casualness (2-tailed t-test,  $p = 0.026$ ) for the 'luggage' sentence while nearly the same for the 'hello' sentence. This difference found between the two sentences may be

due to the difference in the strength of the division between Phrase 1 and Phrase 2 in both sentences: in the 'hello' sentence, Phrase 1 ('hello') is completely independent of Phrase 2 ('is that Mr. Akagi speaking?'), while Phrase 1 of the 'luggage' sentence indicates the subject of the sentence, and thus both phrases are more tightly connected. To summarise the measurements of the last morae in Phrase 1, the durations were consistently longer, (although the difference did not reach the significance level of 0.05), and the steepness of the final fall was significantly more gentle for the 'luggage' sentence for politeness. This difference in  $f_0$  final movement in Phrase 1 appears to be related to the way utterance is divided. So phrasing could be an important factor for distinguishing politeness and casualness.

#### **5.2.6. Differences in voice quality and articulation**

First, an informal auditory assessment was conducted by the author to examine whether any difference in voice quality or articulation was noticed in these two different speaking styles. No obvious difference was perceived in all the speakers except TN and KS for voice quality and TK for articulation. The difference in voice quality appears to be a tense/lax voice difference: the polite utterances of TN and KS for both sentences sounded 'muffled', 'soft' and 'hesitant', while the casual versions 'resonant', 'strong' and 'straight'. The difference in articulation found in TK's utterances was precision: the polite versions were perceived as more 'careful' while the casual versions more 'careless' or 'sloppy'.

Since there were noticeable differences in voice quality and articulation in polite and casual utterances by some speakers (TN, KS and TK), spectral analyses were conducted using spectrograms. The utterances of the first phrase of the 'hello' sentence ('moshimoshi' meaning 'hello') spoken by KS and TK were selected for investigation. The 'hello' phrase was selected because this is the start of the telephone conversation,



and thus very important in terms of conveying a good impression of the speaker. The speakers KS and TK were selected because the polite 'hello' and casual 'hello' of both speakers sounded clearly different, and also because both speakers were judged as very good encoders for politeness and casualness for the 'hello' sentence in the utterance evaluation test with five listener-judges (see Table 4.3).

Figs. 5.1 to 5.4 show the wide and narrow bandwidth spectrograms of the 'hello' phrase ('moshimoshi') spoken by KS and TK in both polite and casual styles. KS's polite utterance shows low frequency emphasis especially in the vocoid segments (i.e., 'mo' and 'i') in contrast to rather high frequency emphasis of its casual counterpart (Figs. 5.1 and 5.2). Since high frequency emphasis suggests that the voice source is stronger, this low/high frequency emphasis difference of the polite and casual utterances agrees with the auditory impression: the polite utterance sounded 'soft' while the casual utterance 'strong'. The upper formants of KS's casual 'mo' are much clearer than those of the polite 'mo'; hence the casual 'mo' can be said to be a more vocoid-like sound whereas the polite 'mo' is more contoid-like (Fig. 5.1). Since vocoid-like sounds are generally associated with a relatively open and unobstructed vocal tract shape, this may be an acoustic manifestation of the 'resonant'/'muffle' difference in voice quality in the auditory assessment. These spectral differences which were found in KS's utterances were not shown clearly in TK's utterances (Fig. 5.3).

The spectrograms (Figs. 5.1 and 5.3) showed a clear difference in the articulation of 'sh' in the utterances of both speakers; the polite 'sh's (especially the first 'sh') are much longer and more clearly defined than their 'casual' counterparts. Voicing continues throughout the first 'sh' in the casual utterances (Figs. 5.2 and 5.4), showing close co-articulation between 'mo' and 'sh', which is a typical sign of fast and/or careless speech. Another interesting difference in articulation was found in KS's last 'i': The polite 'i' maintained the positions of the first three formant very well from the beginning to the



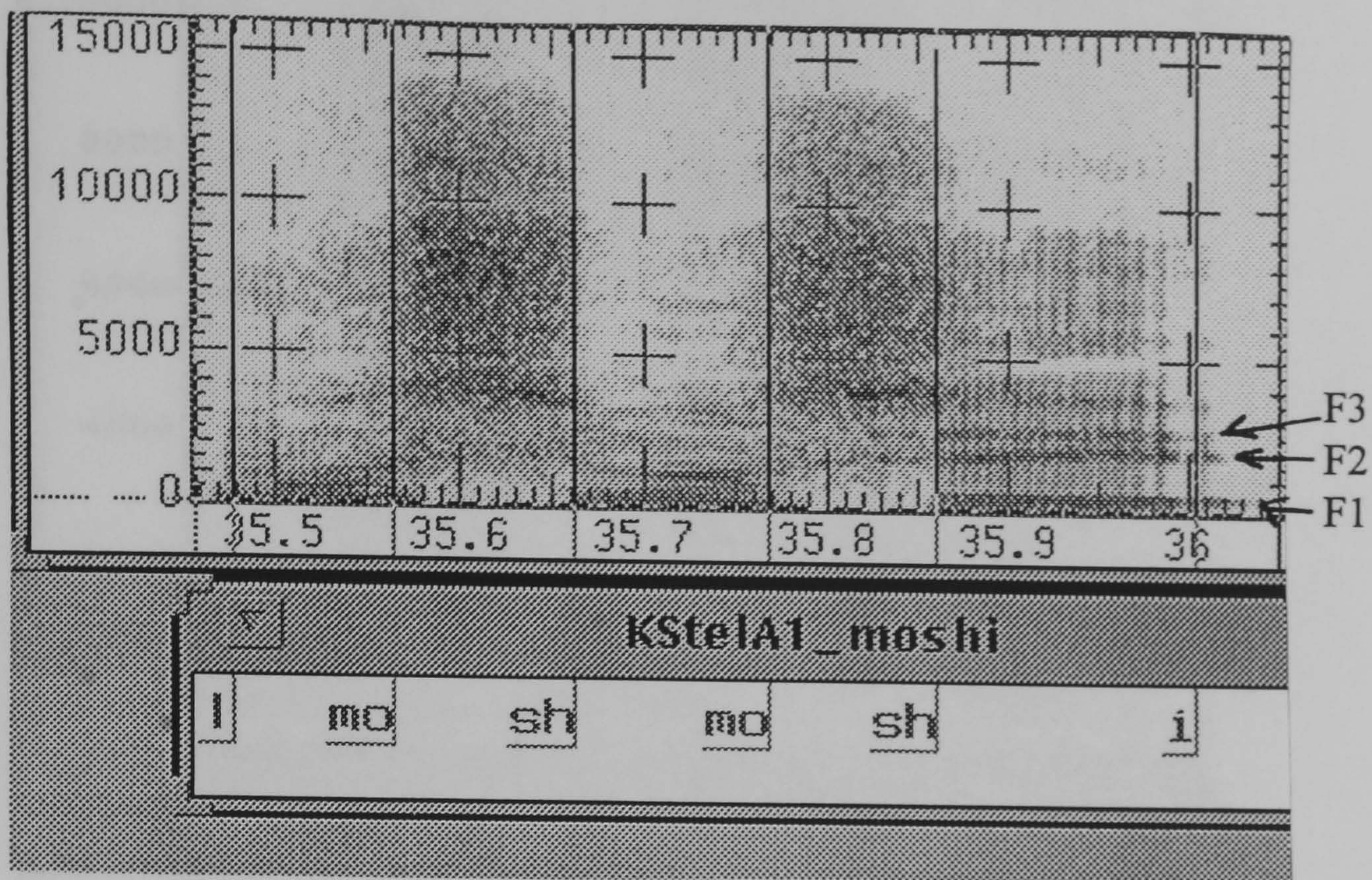
end, while the casual 'i' did change the formant positions: it slightly raised the first formant and lowered the upper formants, so that the sound became closer to a centralised schwa-like sound. Since the speaker needs effort to maintain a clear 'i' sound, because otherwise the tongue and the lips return to the neutral position (which produces a more centralised sound), this could be one of the cues for the careful/careless difference in articulation.

### 5.3. Summary

This chapter described acoustic analyses of polite and casual utterances of two sentences spoken by six male Japanese speakers. In these acoustic analyses, mainly  $f_0$  and temporal variables were focused on. The purpose was to investigate distinct acoustic variables which could be identified as consistently distinguishing different speaking styles. All the temporal variables (i.e., articulation rate, total utterance length and duration of the final morae) were found to be consistently differently used across the six speakers for the polite and casual speaking styles, and therefore, they could be an important cue for politeness. On the other hand, the contribution of the  $f_0$  related variables (i.e., mean  $f_0$ , range, rate of change in  $f_0$ , and  $f_0$  final direction) was inconclusive: great variability was found among the six speakers in use of these  $f_0$  variables in polite and casual utterances. The variability, however, does not necessarily mean that  $f_0$  variables cannot be one of cues for politeness, but it seems unlikely that these  $f_0$  related variables alone play an influential role in perception of politeness. Auditory assessment and spectral analyses of spectrograms were conducted to examine differences in voice quality and articulation. Although many of the speakers did not change their voice quality and the way of articulation in a very noticeable way, differences between these two styles, in terms of energy distribution, co-articulation and the quality of a vowel and a consonant, were found in some speakers' utterances.



(a) POLITE style



(b) CASUAL style

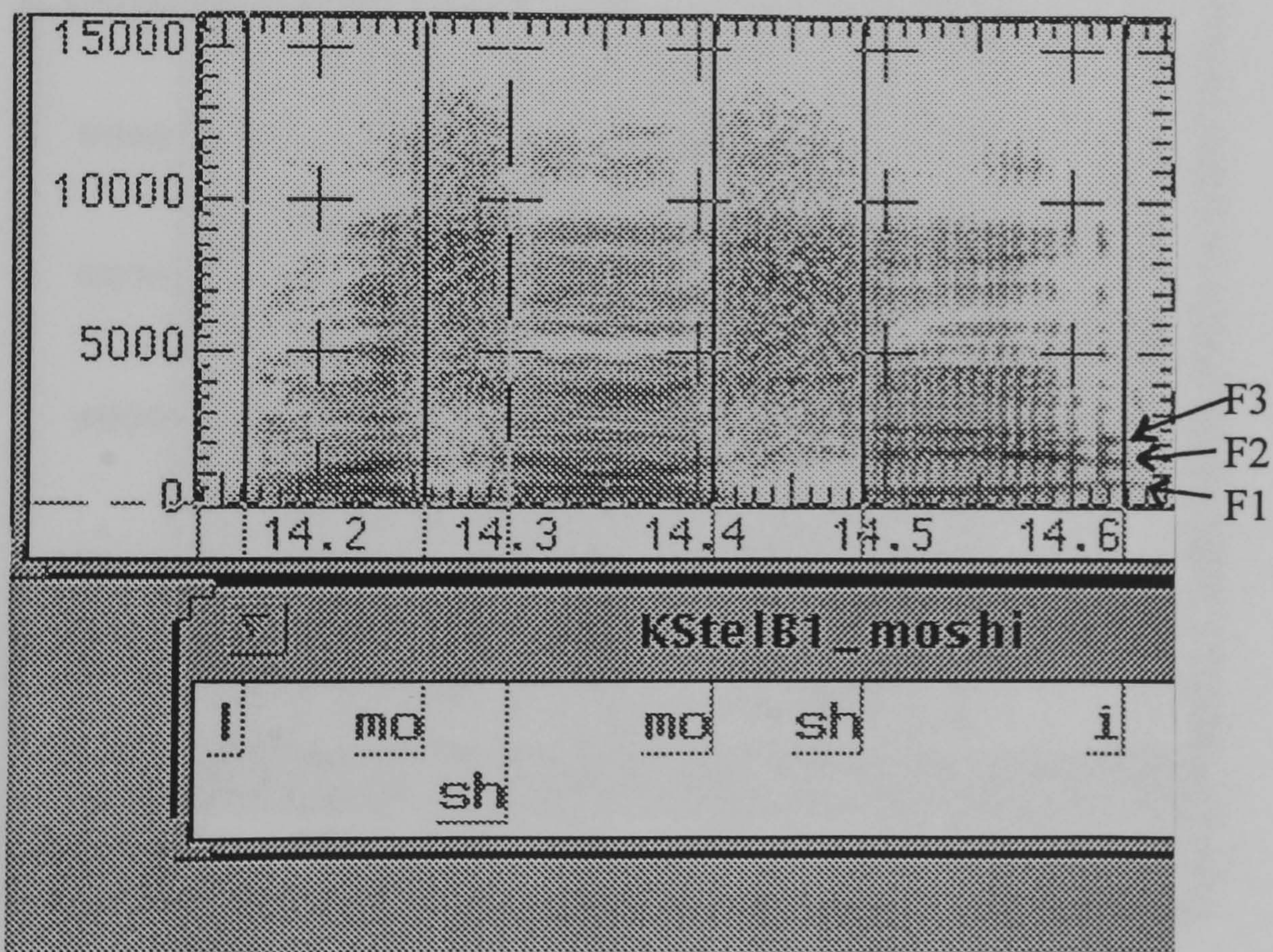
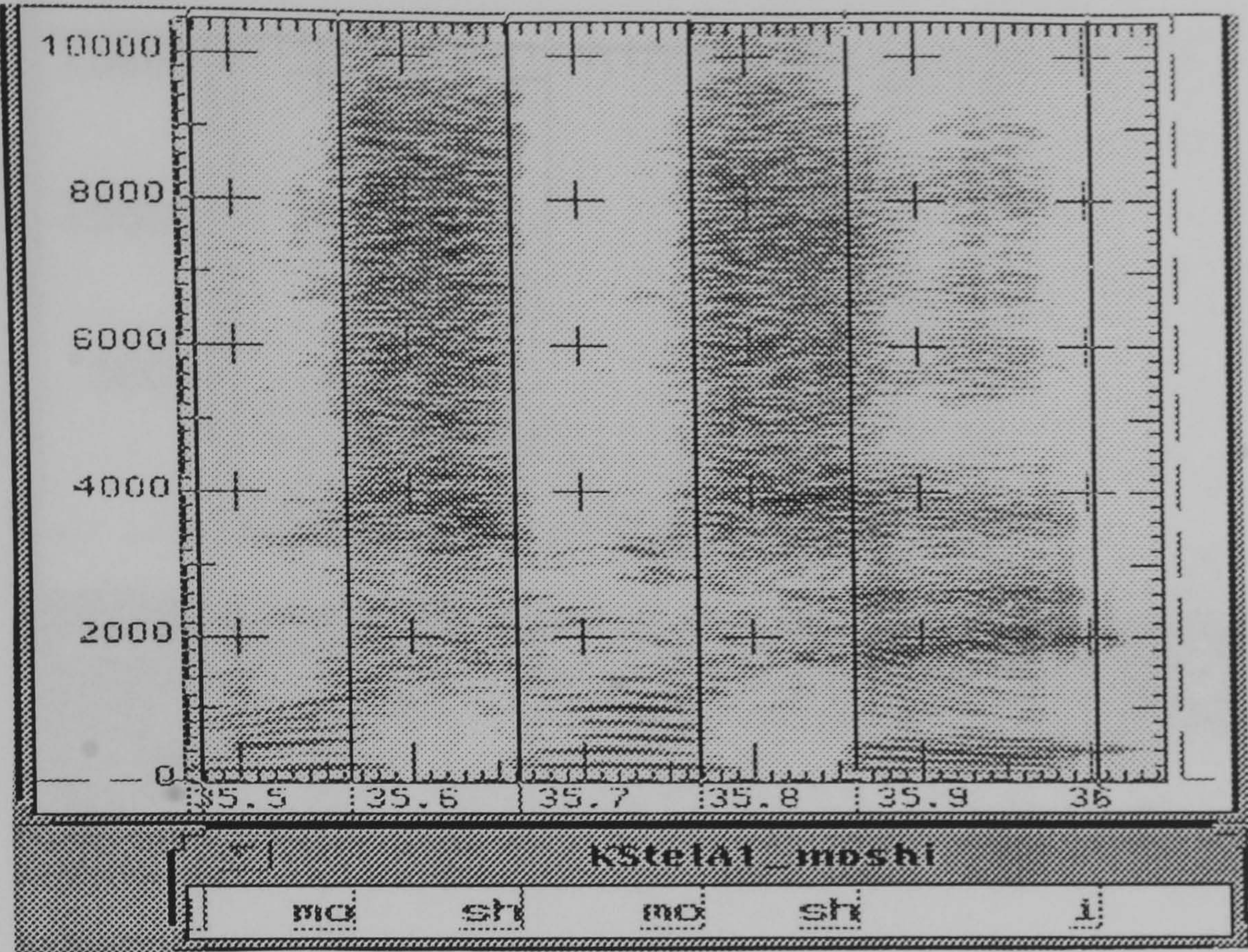


FIG. 5.1. Wide bandwidth spectrograms of 'moshimoshi' spoken by KS. The 'polite' style (a) and the 'casual' style (b) are shown separately.



(a) POLITE style



(b) CASUAL style

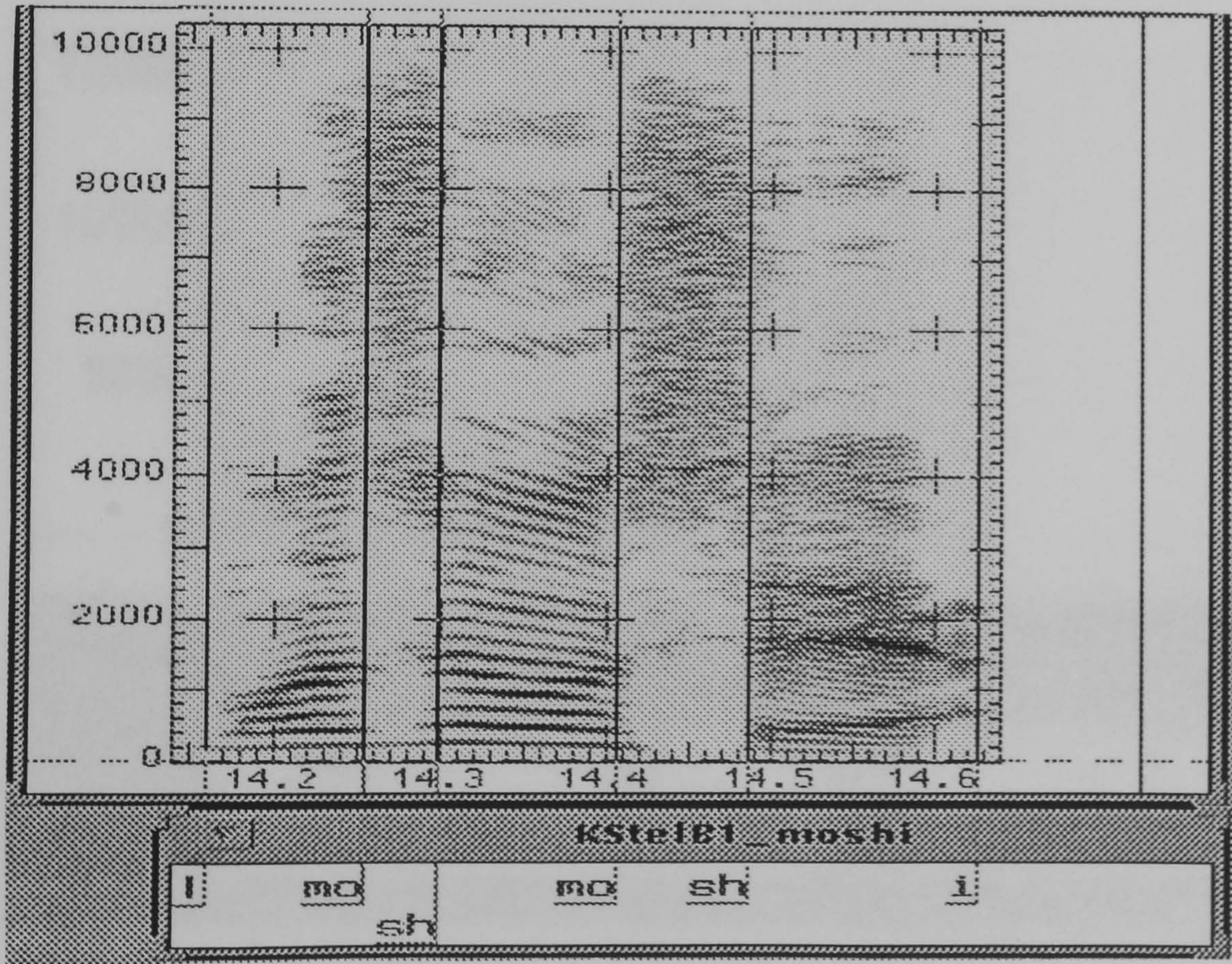
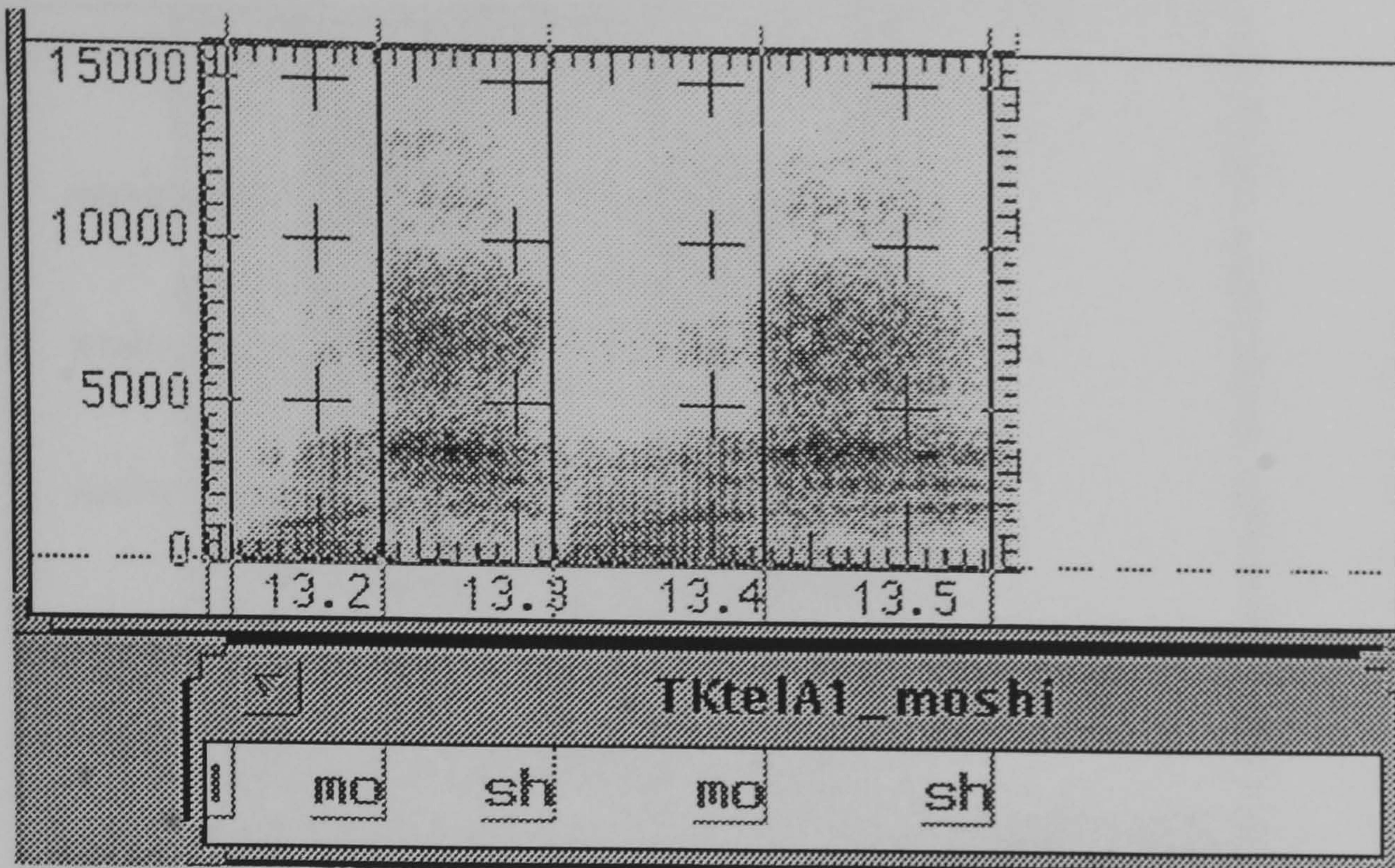


FIG. 5.2. Narrow bandwidth spectrograms of 'moshimoshi' spoken by KS. The 'polite' style (a) and the 'casual' style (b) are shown separately.



(a) POLITE style



(b) CASUAL style

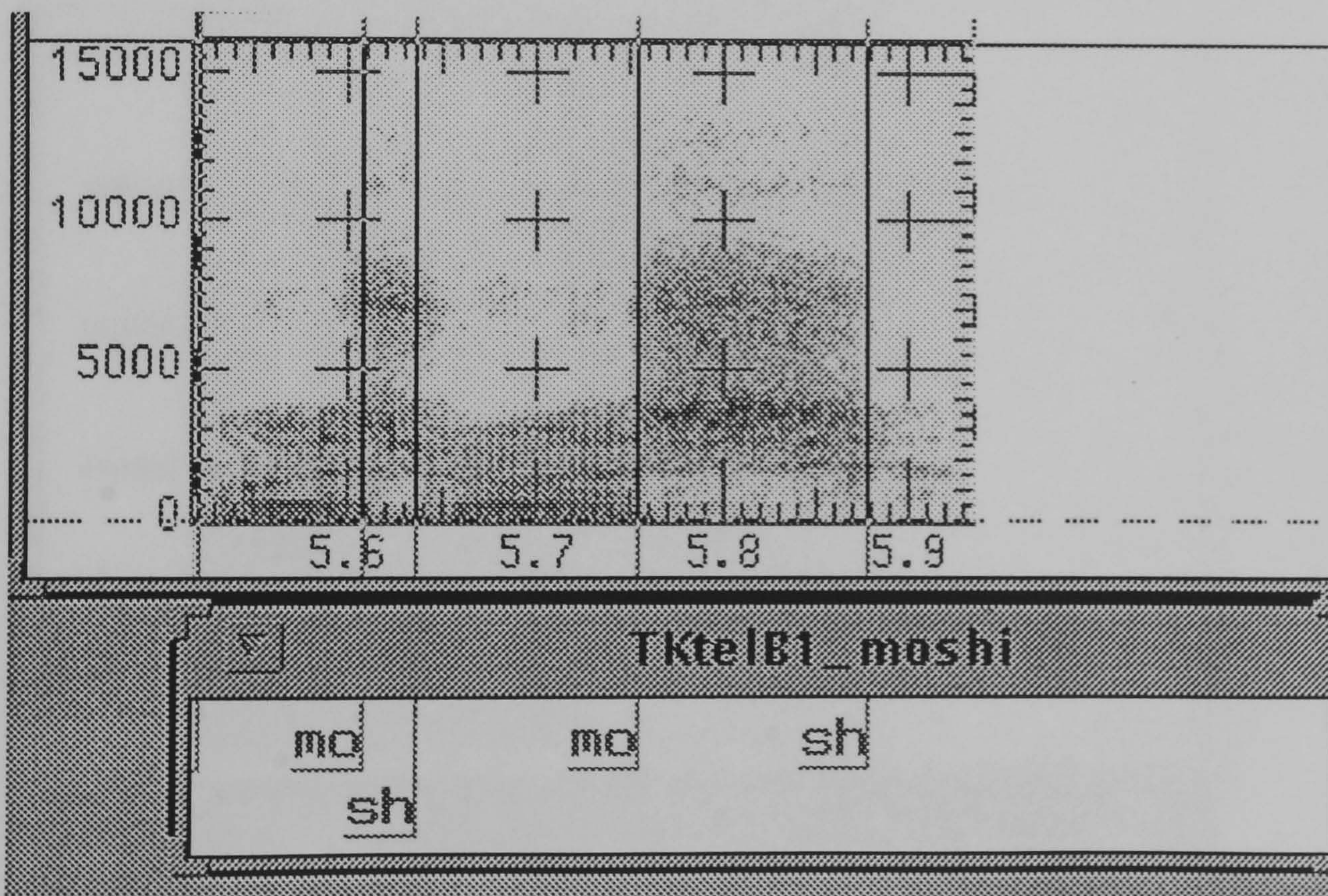
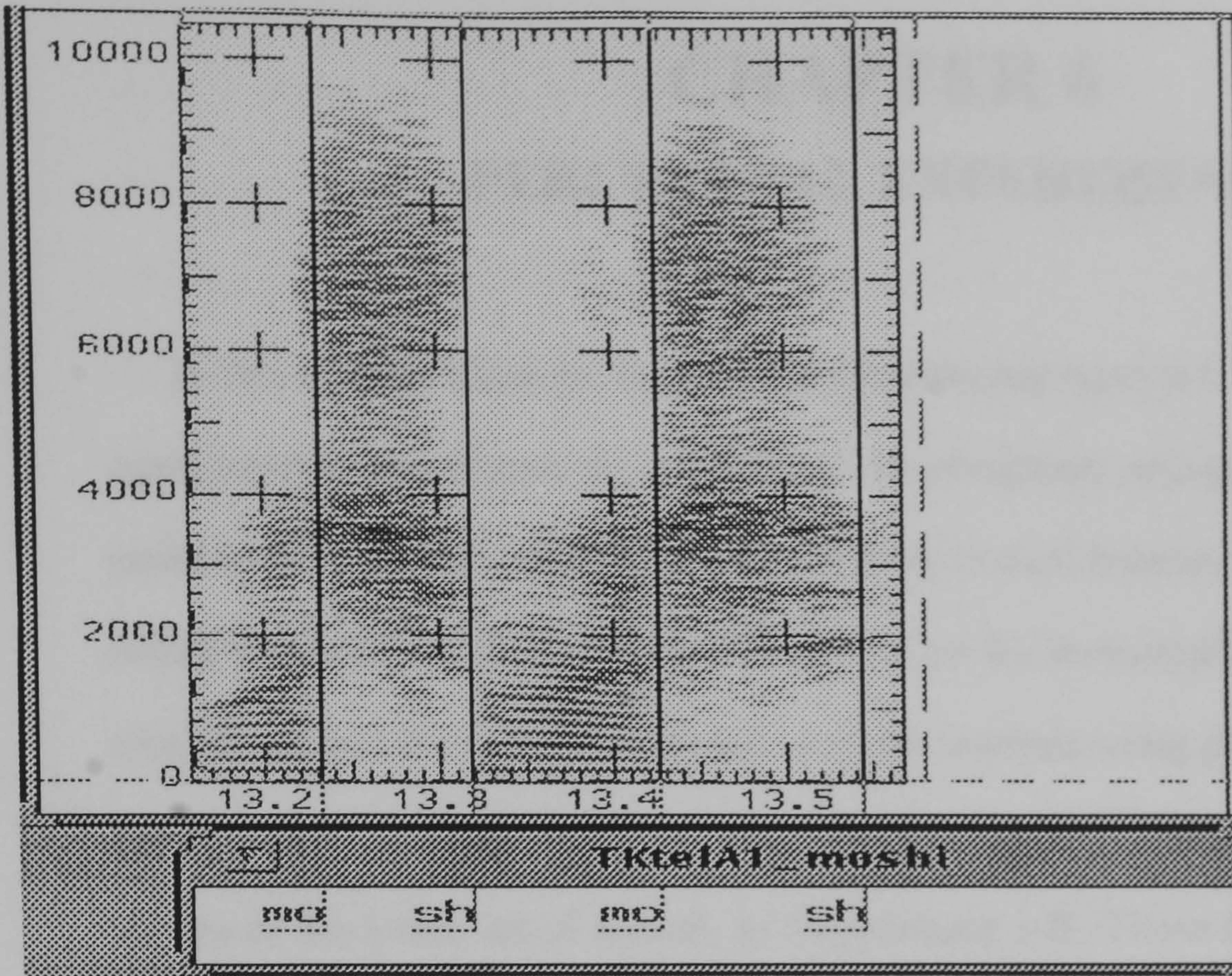


FIG. 5.3. Wide bandwidth spectrograms of 'moshimoshi' spoken by TK. The 'polite' style (a) and the 'casual' style (b) are shown separately.



(a) POLITE style



(b) CASUAL style

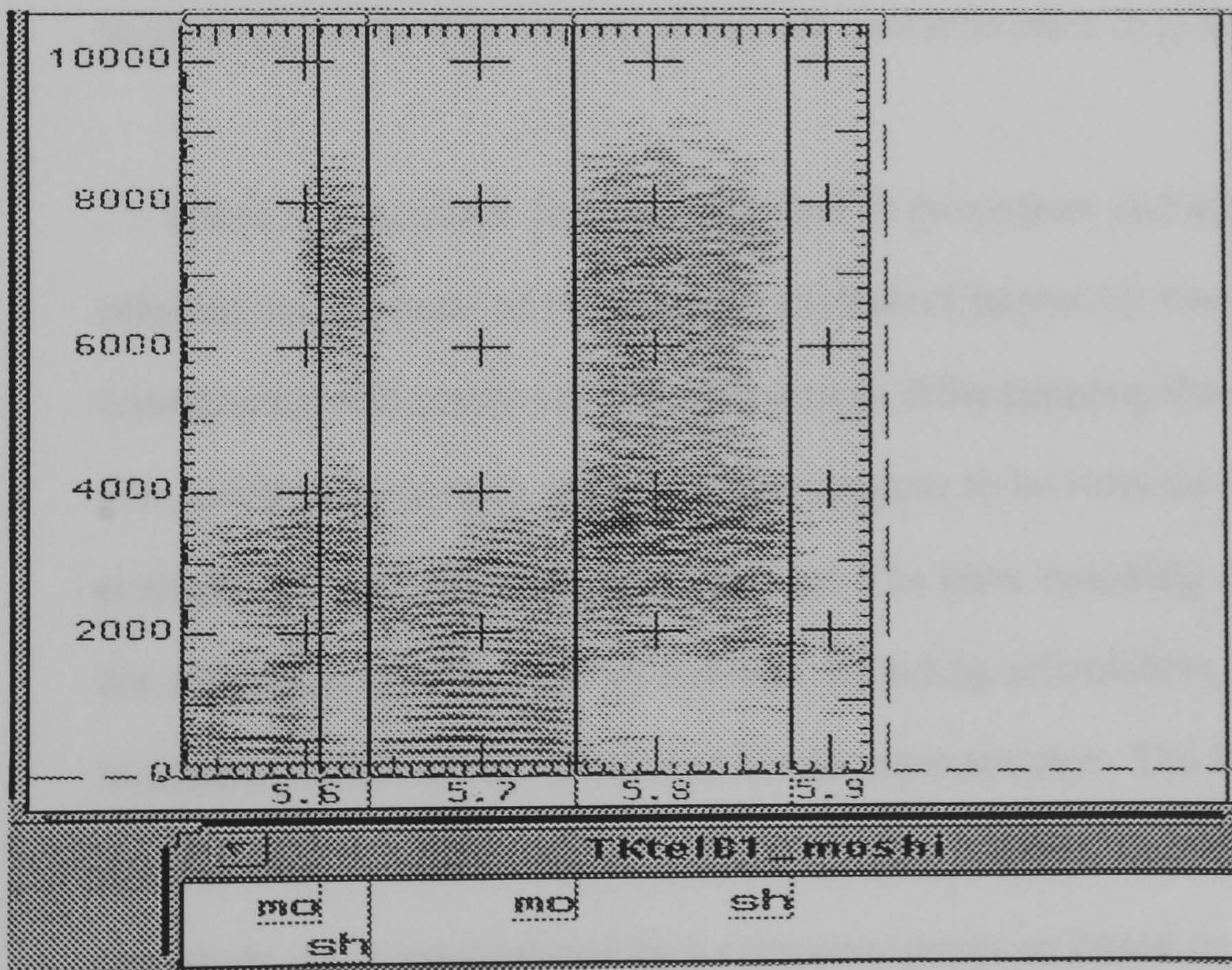


FIG. 5.4. Narrow bandwidth spectrograms of 'moshimoshi' spoken by TK. The 'polite' style (a) and the 'casual' style (b) are shown separately.



## CHAPTER 6

### PERCEPTUAL EXPERIMENTS

In this chapter three perceptual experiments are reported. Experiment 1 was conducted to investigate the role of final  $f_0$  movement and speech rate in signalling politeness. Two sets of stimuli are presented to each listener-judge in the same rating session. The stimuli consisted of one set for investigating the role of final  $f_0$  movement and the other for speech rate. An analysis using politeness scores of the former set of stimuli is referred to as Experiment 1-A, and an analysis using the scores of the latter set of stimuli, as Experiment 1-B. These two analyses are reported in separate sections (Sections 6.1 and 6.2). Experiment 2 again focused on the final part of utterances, but used a different set of utterances of the same sentence used in Experiment 1. A final experiment (Experiment 3) was conducted to investigate the importance of listener characteristics in politeness judgements.

These two acoustic features, the final  $f_0$  movement and speech rate, were selected on the basis of the acoustic analysis (Chapter 5): they were found to be consistently differently used by speakers in differentiating the two speaking styles (i.e., polite and casual), therefore, they appear to be relevant to the degree of politeness conveyed by these two styles. The term 'speaking style' is used as a label for a complex of a number of features, including articulation, pitch, rhythm, loudness, voice quality, produced by the same speaker. The style difference was produced by human speakers' subjective manipulation, and  $f_0$  and segmental durations were manipulated by a computer program based on the TD-PSOLA technique (Charpentier and Stella, 1986).

Utterances of a single sentence (the 'luggage' sentence) spoken by two male speakers were used in Experiments 1 and 3. The reasons why the 'luggage' sentence



and the two speakers were selected are as follows. The 'luggage' sentence was selected because it was a routine question but not as conventional as the 'hello' sentence in the telephone conversations; also utterances of the 'luggage' sentence had more variability in speech rate. In order to determine which speakers out of the six speakers, who took part in the recording sessions, to use, an utterance evaluation test was carried out in terms of politeness. (The procedure and results are described in Section 4.2.) The two of these were selected as the source utterances for the perceptual experiments. The selected speakers, KS and TK, were judged as good speakers, differentiating polite and casual versions clearly (see Table 4.3). Although KS's polite utterance was judged as the least polite among the six speakers' polite versions for the 'luggage' sentence, his casual utterance was rated as the best representative of casualness.

## **6.1. Experiments on the role of the final part of utterances**

These experiments were conducted to investigate the effects of final  $f_0$  movement of the final vowel on politeness judgements. Experiment 1 was conducted using utterances by two untrained male speakers, KS and TK. Experiment 2 used utterances spoken by one trained male speaker.

### **6.1.1. Experiment 1-A: The effects of final $f_0$ movement**

#### **6.1.1.1. Method**

##### *1. Design*

A factorial  $2 \times 2 \times 2 \times 2$  design was used with two speakers, two types of styles (polite and casual), two levels of duration of the final vowel (short and long), and two types of  $f_0$  direction (rise and fall).

## 2. *Speech materials and stimulus preparation*

The polite and casual versions of the 'luggage' sentence spoken by two speakers (KS, who came from Kyushu, the southern island, and TK from Yokohama, the eastern part of Japan) were used as source utterances. F0 and duration of the vowel in the final mora 'ka' in 'Nimotsu ... KA', a sentence final particle indicating that the sentence is a question, were measured in order to determine the values used for the following cue manipulation. Eight patterns were resynthesised based on the polite/casual version of the source utterance spoken by each speaker by manipulating f0 and duration of the final vowel 'a'. The three factors were: two styles ('polite' and 'casual'), two durations of the final vowel (short and long) and two types of f0 movement of the final vowel (rise and fall).

The final vowel of the 'polite' version of Speaker KS had an initial f0 point of 110 Hz and decreased by 54 semitones per sec over 115 ms, while the TK final vowel had an initial f0 point of 110 Hz and increased by 28 semitones per sec over 70 ms. The KS's final vowel of the 'casual' version started from 117 Hz and slightly decreased 9 semitones per sec over two thirds of the total duration of 240 ms, and then increased at the rate of 56 semitones per sec over the rest; TK's final vowel started from 101 Hz and increased by 20 semitones per sec over 130 ms.

The values for the duration factor were set to 120 ms for KS and 70 ms for TK, for the 'short' duration, and 240 ms for KS and 130 ms for TK, for the 'long' duration. For the f0 contour, a straight line was used for all the versions, except the long versions by KS, for which two straight lines were used (that is, a slightly falling one and a rising one) for the sake of naturalness. The acoustic analysis showed that the rate of change of the final vowel in the source utterances was between 20 and 30 semitones per sec for TK and between 50 and 60 semitones per sec for KS, but the latter very steep fall/rise versions were judged somewhat less natural in a pilot



study, when they were resynthesised. Therefore,  $\pm 25$  semitones per sec was adopted as the rate of change with the initial  $f_0$  points of the source utterances.

There were six occurrences for each stimulus condition. Stimuli were mixed with the stimuli for the experiment on the speech rate (Experiment 1-B) in random order. A total of 164 stimuli, consisting of six occurrences for a total of 26 conditions (16 conditions for this experiment and 10 conditions for Experiment 1-B) and a total of eight dummy stimuli at the beginning and the end of three sub-sessions, were recorded on high-quality audio cassette tapes in random order. Each utterance was preceded by a warning tone and followed by a three-second silence during which subjects were asked to make ratings.

### 3. Rating scale

The bipolar 8-cm scale of politeness shown in Fig. 6.1 was used.

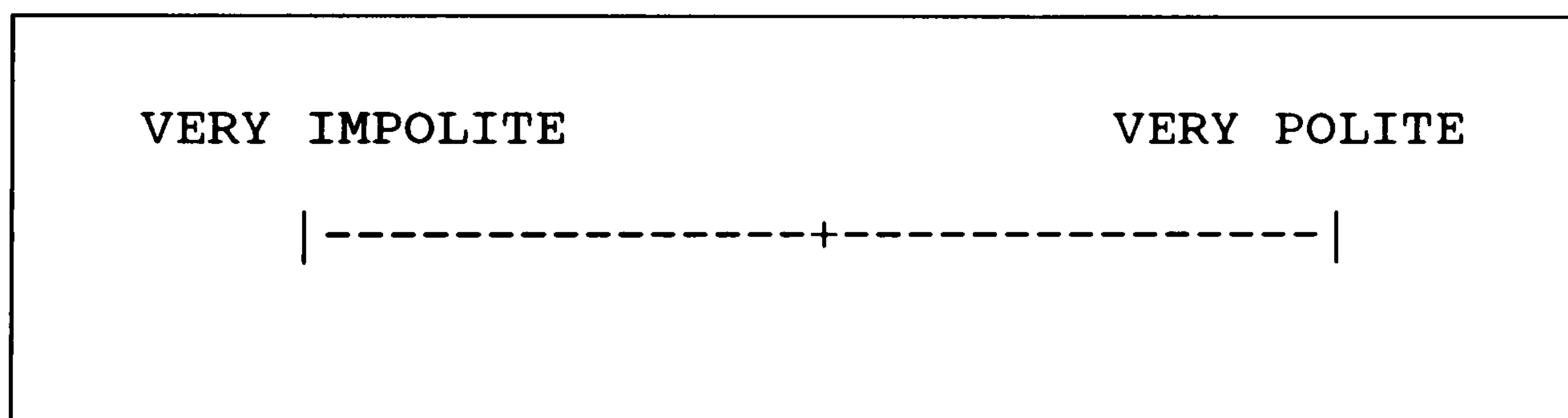


FIG. 6.1. Rating scale of politeness used in Experiment 1.

### 4. Subjects

Twenty paid subjects (12 male and 8 female), mostly students from two universities in the eastern part of Japan, participated in the experiment. They were all native speakers of Japanese, ranging in age between 20 and 36. Seven male and five female subjects were from the eastern part of Japan, four male and two female

from the western part, and one male and one female from other parts of Japan.

### *5. Rating sessions*

Subjects were given written instructions in Japanese, telling them that they would hear only one Japanese sentence spoken in various ways, and their task was to rate utterances on a scale of politeness according to their own evaluation criteria. They were also informed that the speakers were young customs officers trying to speak politely to a respectable gentleman and also rather casually to a young student, and that all the utterances they were going to hear would sound neither very polite nor very impolite.

At the beginning of the session, the subjects were presented with polite and casual source utterances by these two speakers in order for them to assess the range of politeness. In a practice session they listened to six stimuli including the four source utterances and two utterances with the maximum degrees of manipulation used in this experiment and rated them on the politeness scale. They then listened to the 164 stimuli in random order over Sennheiser HD 480II headphones in three sub-sessions with a short break between them. At the end of the session the subjects' speech rate was assessed; this procedure is described in Experiment 1-B (Section 6.2). The 20 subjects were tested individually in a quiet room, each session lasting about 40 minutes. The instructions given to the subjects are attached in Appendix E.

#### **6.1.1.2. Results and discussion**

The politeness scores were obtained by measuring the distance between subjects' markings on the linear scale and a mid point on the scale. Scores could range between -4 (very impolite) and +4 (very polite). Kendall's coefficient of



concordance ( $W$ ) was calculated to assess the general agreement among 20 listener-judges' ratings of the four conditions (i.e., the combinations of short and long final vowel duration, and rise and fall final vowel  $f_0$  directions) of both (1) the polite style and (2) the casual style originally spoken by the two speakers, KS and TK. The mean reliability assessed by the Spearman rank-order correlation coefficient (Mean  $r_s$ ) is also reported together with  $W$ : (1) for the polite versions,  $W = 0.53$  (Mean  $r_s = 0.51$ ,  $N = 20$ ) for KS and  $0.57$  (Mean  $r_s = 0.55$ ,  $N = 20$ ) for TK ( $p_s < 0.0001$ ); (2) for the casual versions,  $W = 0.12$  (Mean  $r_s = 0.07$ ,  $N = 20$ ) for both speakers ( $p_s < 0.05$ ). The ratings for the polite style were more consistent than those for the casual style. The effective reliability ( $R$ ) was then calculated to assess the reliability of the ratings of all judges, by using the Spearman-Brown formula (Section 3.5.2.5). The results showed that the effective reliability was very high ( $R = 0.95$ ) for the polite style, and reasonably high ( $R = 0.60$ ) for the casual style of both speakers, and this justifies using rating scores to investigate the effects of cue manipulated stimuli.

The intra-judge agreement was assessed by the ratio between variance of each judge's repetition scores and the total variance (this is described more fully in Section 3.5.2.5). Four ANOVA tests were performed separately on scores for polite and casual utterances by each speaker, with factors of final duration (two durations), final  $f_0$  direction (two directions) and each subject's repetition factor (six repetitions). These tests established that the repetition factor was not significant in all cases except KS's polite utterances ( $p = 0.01$ ). Therefore it is justifiable to use mean values over the six repetition scores in further analyses.

An ANOVA was then carried out with factors of speaker (two speakers), speaking styles (polite or casual), final vowel duration (long or short) and final  $f_0$  direction (rise or fall). The result showed significant main effects of speaker, final vowel duration and final  $f_0$  direction, and significant interactions between speaker

and style, final duration and speaker, and final duration and style ( $ps < 0.05$ ). The results of this ANOVA test are attached in Appendix 2 (1-A-1), and the significant effects at the level of 0.05 or better are summarised in Table 6.1. No interaction between duration and f0 direction of the final vowel was found, which suggests that these two acoustic variables function independently. To assess the relative magnitude of the effects of each factor, eta-squared was calculated. The eta-squared is a statistic based upon the ratio of the sums of squares of a factor to the sums of squares of the total of all the within-subject factors. According to this indicator, the factor of final vowel duration was the most salient factor, and the f0 direction factor was less important. The importance of the temporal aspects of the final part of sentences has been mentioned in Imaizumi *et al.* (1994). They studied the effects of temporal variables in relation to emotions from the point of view of listener-adaptive characteristics in dialogue, and found that the length of the final part of the target sentences 'desuka' played an important role in accounting for a factor which represented the emotional contrast between discomfort (e.g., awful, rough, etc.) and comfort (e.g., easy, kind, polite, etc.).

TABLE 6.1. ANOVA results of Experiment 1-A: significant effects at the level of 0.05 or better.

	<i>F</i>	<i>p: significance of F</i>	<i>eta-squared (%)</i>
<u>MAIN EFFECTS</u>			
Speaker	7.48		8.2
Final Duration	34.73	**	11.2
Final F0 Direction	9.38	*	1.2
<u>INTERACTIONS</u>			
Speaker and Style	11.02	*	3.3
Speaker and Final Duration	10.22	*	1.4
Style and Final Duration	25.54	**	1.4

$df_{effect} = 1, df_{error} = 19$   
\* :  $p < 0.01$   
\*\* :  $p < 0.001$



The mean values of the politeness scores for each condition are shown in Table 6.2. The mean values for the short and long final duration versions for the polite and casual styles by the two speakers are shown in Fig. 6.2, which shows that short duration was rated more positively than the longer duration. The differences between short and long final duration versions were significant in three cases out of four (i.e., polite and casual versions by both speakers) (2-tailed t-tests,  $ps < 0.005$ ), except TK's casual versions.

TABLE 6.2. Mean politeness ratings with standard deviations (SD) in Experiment 1-A.

<i>Speaker</i>	<i>Condition</i>			<i>Mean</i>	<i>SD</i>
	<i>Style</i>	<i>Final duration</i>	<i>Final direction</i>		
KS	Polite	Short	Rise	0.49	1.035
	Polite	Short	Fall	0.18	1.241
	Polite	Long	Rise	-0.97	1.544
	Polite	Long	Fall	-1.27	1.464
KS	Casual	Short	Rise	0.36	1.837
	Casual	Short	Fall	0.31	1.732
	Casual	Long	Rise	-0.63	1.874
	Casual	Long	Fall	-0.73	1.755
TK	Polite	Short	Rise	1.53	1.028
	Polite	Short	Fall	1.23	0.970
	Polite	Long	Rise	0.68	1.158
	Polite	Long	Fall	0.06	0.988
TK	Casual	Short	Rise	0.45	1.143
	Casual	Short	Fall	-0.06	1.149
	Casual	Long	Rise	0.12	1.093
	Casual	Long	Fall	-0.06	1.297

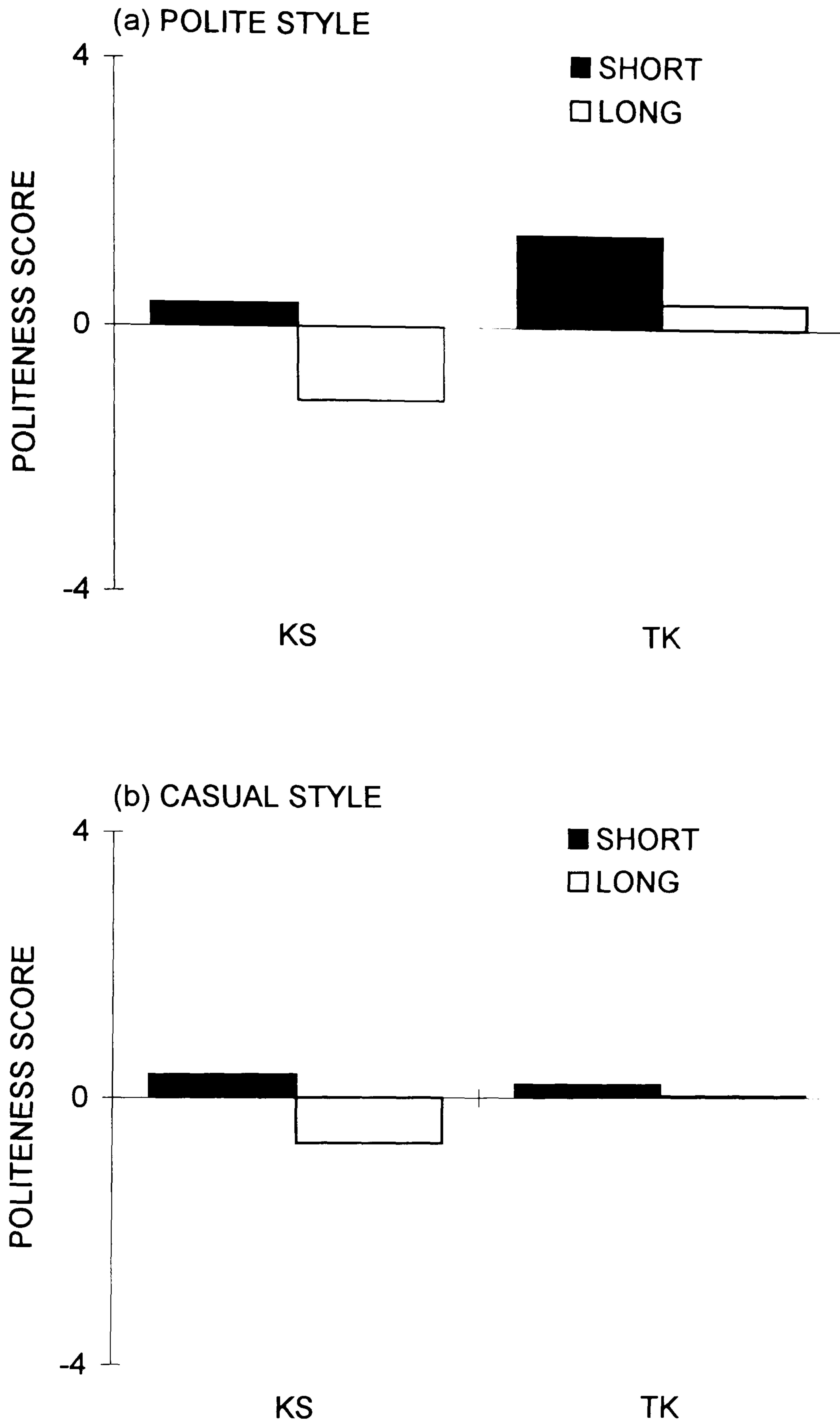


FIG. 6.2. Comparisons between the mean values of politeness scores for the short final duration (SHORT) and long final duration (LONG) versions of the utterances originally spoken by KS and TK in Experiment 1-A; the scores for the polite style (a) and those for the casual style (b) are shown separately.



The mean values of the politeness scores for the final rise and final fall versions for the polite and casual styles by the two speakers are shown in Fig. 6.3, which shows that final rise was rated as more polite than the final fall. However, the differences were less salient in comparison to those in the final duration of the utterance. Although the differences of mean values between the final rise and final fall versions are not very large, the majority of raters did prefer the final rise version to the final fall version in their politeness judgements: 52.5% of the subjects rated the final rise more positively than the final fall, while only 23.8% of the subjects preferred the final fall and 23.8% of the subjects showed no preference.

This final rise preference in relation to politeness may be related to unmarkedness of sentence intonation, because the sentence used is a direct yes-no question whose unmarked intonation is a rising tone. This result agrees with the finding of Scherer *et al.* (1984), who examined the sentence final intonation of German sentences in relation to several attitudinal meanings. They found that a final fall was rated more positively in wh-question sentences, while a final rise had higher ratings in yes-no question sentences, on scales of agreeability and politeness. Since the unmarked intonation of wh-questions is a falling tone while that of yes-no questions is a rising one, they suggested that their results reflected the preference for the traditional description of 'normal' or 'unmarked' intonation. Similarly, Ogino and Hong (1992) studied the sentence final intonation in Japanese in relation to politeness and found that a level or a falling tone was identified with polite versions for an expression 'deshouka', whose default tone is level or a slight fall (although no clear pattern was seen for non-polite utterances).

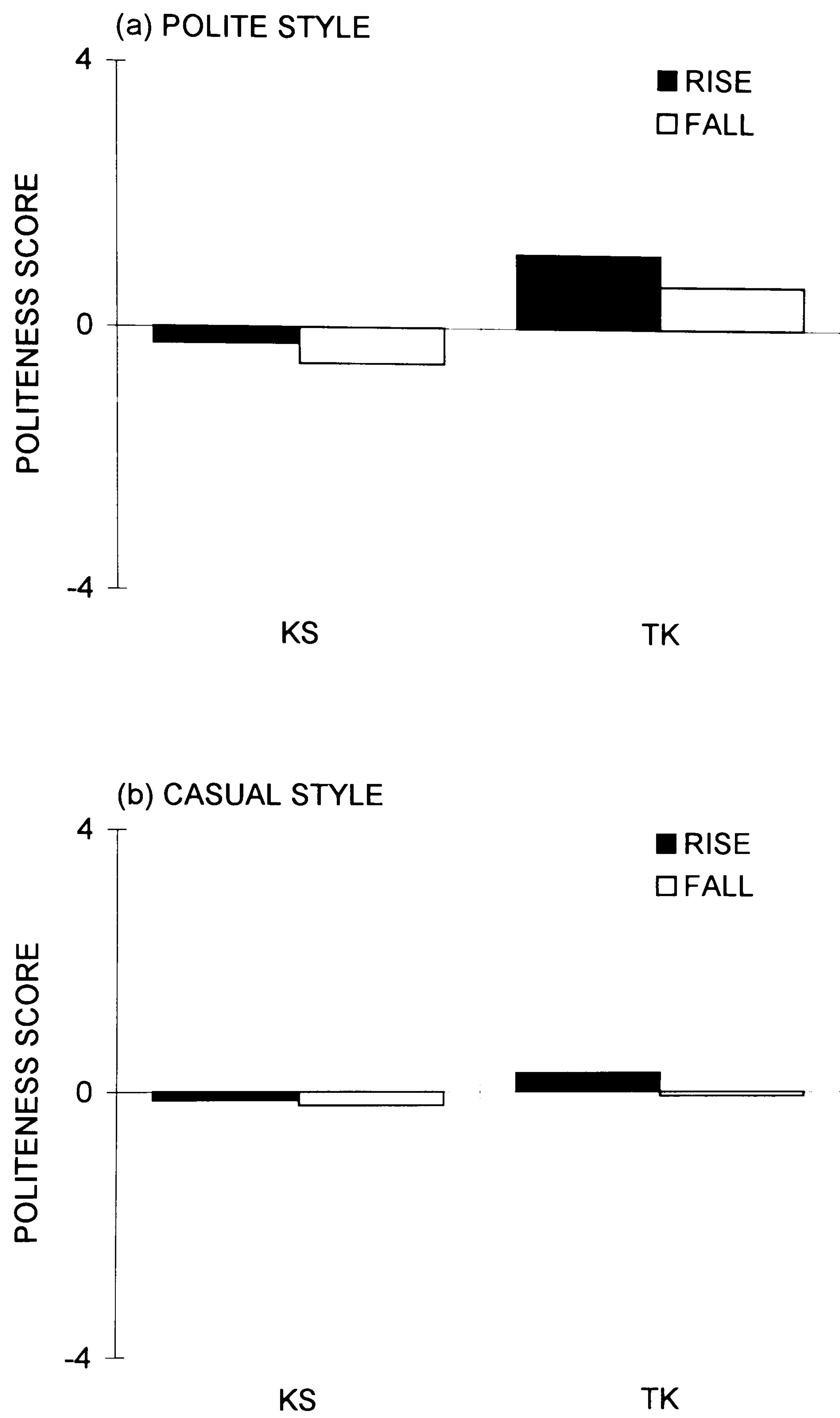


FIG. 6.3. Comparisons between the mean values of politeness scores for the final rise (RISE) and final fall (FALL) versions of the utterances originally spoken by KS and TK in Experiment 1-A; the scores for the polite style (a) and those for the casual style (b) are shown separately.



Fig. 6.4 shows the mean politeness ratings for the four conditions for each speaker (KS and TK): the speaker's polite style with a 'preferred' final prosody (i.e., short duration and a rising tone of the final vowel), which is referred as 'Matched final prosody' in the figure; and also with a 'less preferred' final prosody (i.e., long duration and a falling tone of the final vowel), which is referred to as 'Conflict final prosody'<sup>1</sup>. Also shown is the speaker's casual style with a preferred final prosody ('Conflict final prosody') and with a less preferred final prosody ('Matched final prosody'). For both speakers' utterances, the polite style with the preferred final prosody was rated more polite than their casual style with the less preferred final prosody (1-tailed t-tests,  $ps < 0.005$ ) The casual style with the preferred final prosody was rated significantly more polite than the polite style with the 'less preferred' final prosody for KS's utterances ( $p < 0.005$ ), and the polite style with the 'less preferred' final prosody did not sound more polite than the casual style with the 'preferred' final prosody for TK's utterances. This pattern of results shows that subjects were heavily influenced by the sentence final prosody when they made politeness judgements.

---

<sup>1</sup>: terms 'matched' and 'conflict' are used in the sense that 'polite' style and 'preferred' final prosody, and 'casual' style and 'less preferred' final prosody, are 'matched', whereas 'polite'/'casual' style and 'less preferred'/'preferred' final prosody are 'in conflict'.

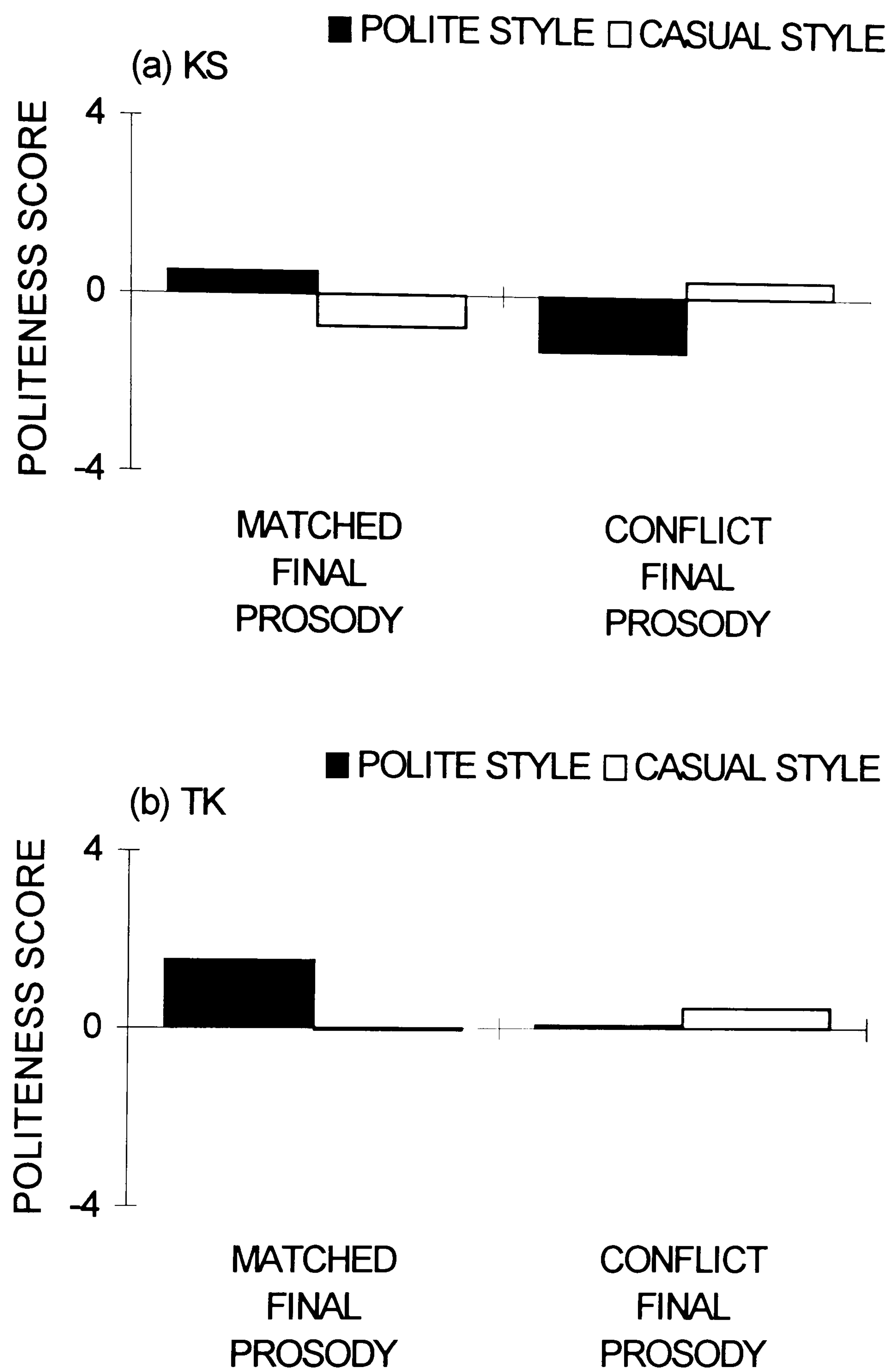


FIG. 6.4. Mean values of politeness scores rated by 20 subjects for the polite/casual source utterances with the 'matched' and 'conflict' final prosody in Experiment 1-A. The scores for the utterances by KS (a) and those for the utterances by TK (b) are shown separately.



Since the accent of the eastern part of Japan and that of the western part of Japan are different in various ways including intonation and articulation, the effects of raters' accent in relation to politeness judgements were also examined. The subjects were categorised into three accent groups: (1) eastern, (2) western and (3) others. An ANOVA test was performed with factors of speaker, style, final duration and final  $f_0$  direction as within-subjects factors, and accent of the subjects as a between-subjects factor. The results are attached in Appendix 2 (1-A-2). The test showed a significant main effect of final duration, and significant interactions between accent and style, speaker and style, and style and final duration ( $ps < 0.05$ ). The main effect of final  $f_0$  direction and the interaction between accent and speaker just failed to reach significance ( $p = 0.06$ ). The mean ratings for the polite style and the casual style spoken by an eastern speaker, TK, rated by the subjects from the eastern part of Japan and those rated by the subjects from the western part of Japan are shown in Fig. 6.5 (b). A similar pattern was obtained for the utterances by a southern speaker, KS, except that the eastern raters rated KS's polite versions much less polite (Fig. 6.5 (a)). It is interesting to observe that the eastern raters seem to have perceived the eastern speaker's intention correctly, by rating TK's polite utterances polite and his casual versions less polite, while the western raters failed to do so. There was no significant interaction between the accent factor and the factors of any acoustic variables studied here (i.e., the final duration and final  $f_0$  direction). This suggests that the politeness ratings associated with these acoustic variables were not affected by the subjects' own accent. In other words, subjects showed difference in style preference, whereas they responded to the differences in these acoustic variables in the same way.

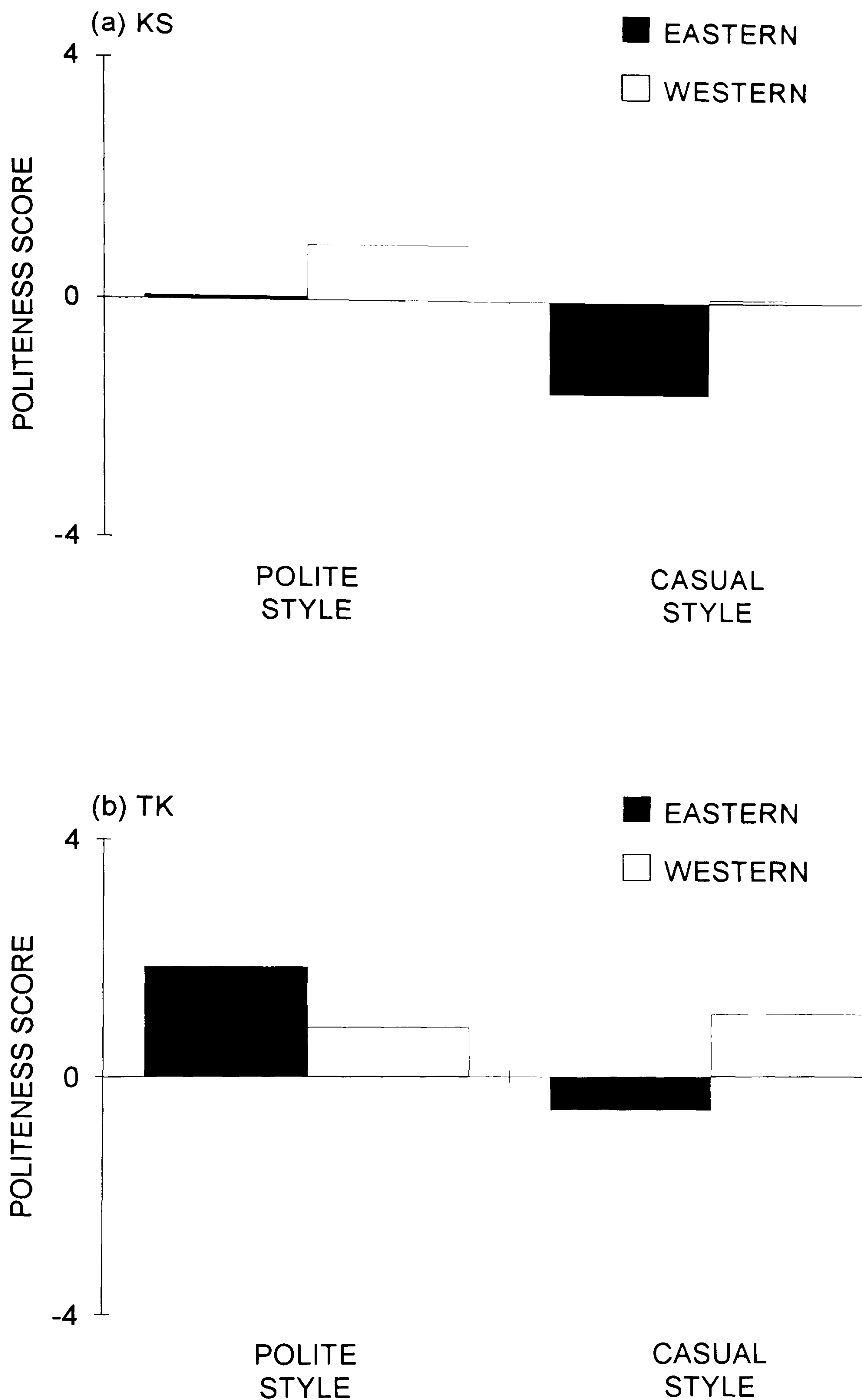


FIG. 6.5. Style preferences of 20 subjects for polite (i.e., the polite style with the 'preferred' final prosody) and casual (i.e., the casual style with the 'less preferred' final prosody) utterances spoken by two speakers, according to the accent of the subjects in Experiment 1-A. The scores for the utterances by KS (a) and those for the utterances by TK (b) are shown separately.



### 6.1.2. Experiment 2: The effects of speaking style of the final part

This experiment was conducted to examine the importance of the final part in a new set of utterances (i.e., utterances spoken by a highly trained speaker, who adopted clearly different 'voices' for different styles), specifically the effects of different speaking styles<sup>2</sup>, especially those of the final mora, in signalling speaker variables. Utterances of the same sentence used in the Experiment 1-A (i.e., the 'luggage' sentence) spoken by a trained male Japanese speaker who was instructed to speak them angrily and kindly were used as source utterances. Although it has been noted that trained speakers tend to adopt theatrical voices rather than their natural voices, utterances by a trained speaker were used. This is because one of the aims of this experiment was to examine the effects of speaking style, and these utterances used in this experiment had a clear difference between the angry style and the kind style, although both styles might be slightly exaggerated.

#### 6.1.2.1. Method

##### *1. Design*

A factorial 2 x 2 x 2 design was used with two speaking styles of the utterance except the final mora (angry and kind), two speaking styles of the final mora (angry and kind) and two types of f0 movement of the final mora (angry and kind).

##### *2. Stimulus presentation*

Utterances of the 'luggage' sentence spoken by a trained male middle-aged Japanese speaker, who was a professional broadcaster, were used. The speaker was given a short dialogue between a customs officer and a passenger, and instructed to speak the

---

<sup>2</sup>: 'style' is used here to include both politeness and affect variables.

lines of the officer to a person who read the lines of the passenger, in several different ways. The utterances were recorded at ATR Interpreting Telecommunications Research Laboratories in Japan. The details of the recordings are described in Miyatake and Sagisaka (1990). Among those different speaking styles, utterances spoken in an angry/irritated way and a kind/considerate way were used in this experiment. The recordings were later digitised at a sampling rate of 16 kHz onto a Sun workstation. Since the speaker's actual 'kind' utterance sounded unnaturally slow to most native speakers who participated in an informal listening test, segmental durations were linearly compressed by 20% and this compressed version was used as the 'kind' source utterance. The waveforms and f0 contours of the 'angry' and 'kind' source utterances are attached in Appendix F.

F0 and duration of the final mora 'ka' of both speaking styles were measured by using ESPS/Waves. Eight patterns were resynthesised by a computer program based on the TD-PSOLA technique as follows. The source utterance (either 'angry' or 'kind') was decomposed into two parts: the first part "Nimotsu ... desu" and 'ka'. The final mora in either the 'angry' or 'kind' utterance was manipulated in such a way that the duration and f0 values were to realise the actual values of the two types of source utterances. This created four types of final mora (i.e. 'angry' style with 'angry'/'kind' prosody; 'kind' style with 'angry'/'kind' prosody). Finally, the two types of the first part were combined with the four types of the final mora. Schematic figures of the f0 movements of the final vowel 'a' in both the 'angry' and 'kind' source utterance are shown in Fig. 6.6. The figures of 3D plot of the final mora in both the 'angry' and 'kind' styles are also attached in Appendix F.



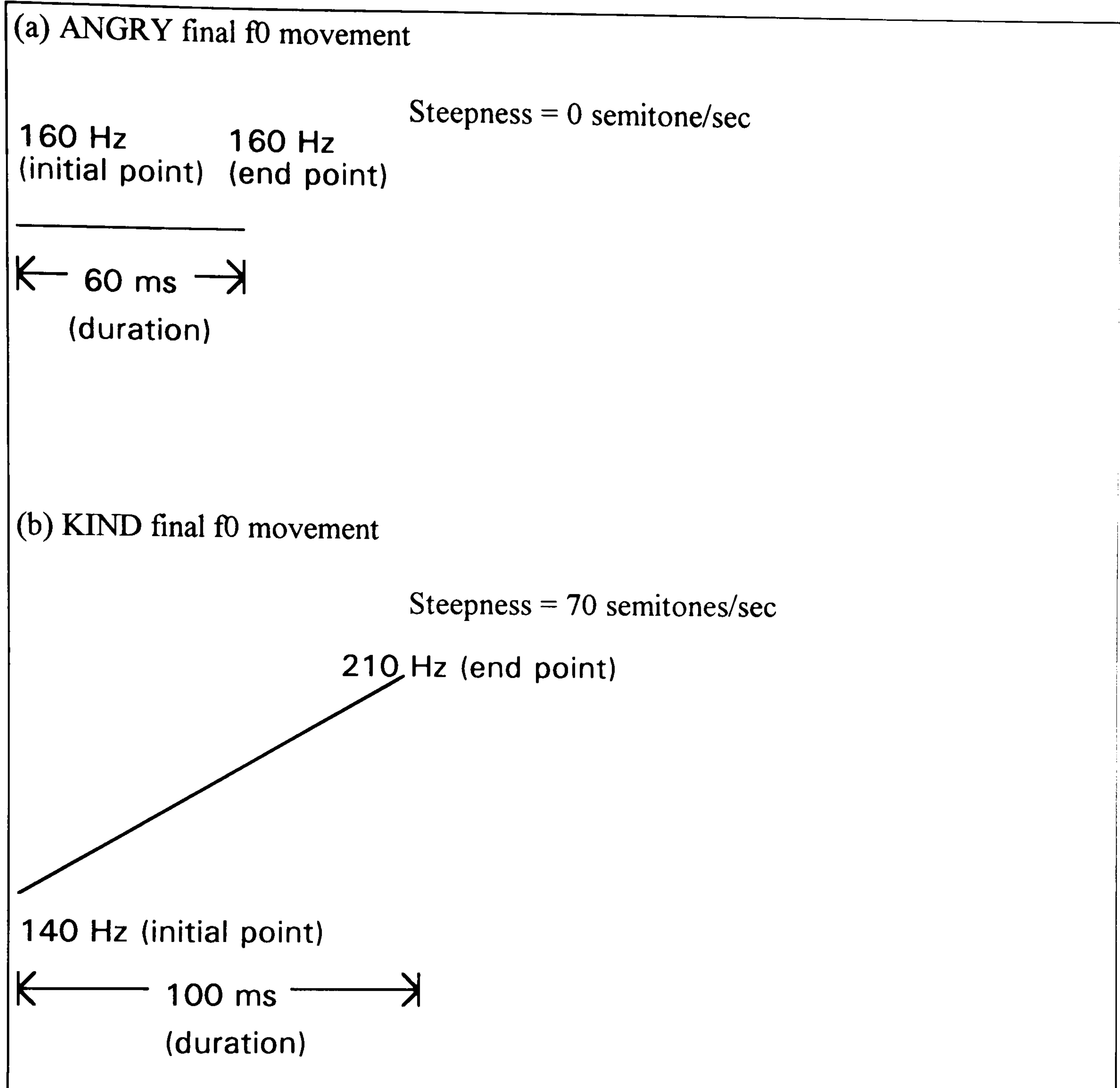


FIG. 6.6. Schematic figures of the actual final f<sub>0</sub> movements. The final f<sub>0</sub> movement in the 'angry' style (a) and the 'kind' style (b) are shown separately.

Since all the f<sub>0</sub> movements were nearly straight, linear interpolation with only one line was used for calculating f<sub>0</sub> values for resynthesis. In order to specify the line fitted to the actual f<sub>0</sub> movement, four factors were to be determined: duration, initial point of f<sub>0</sub>, f<sub>0</sub> direction and rate of change in f<sub>0</sub> (see Fig. 6.6). Noticeable differences between the 'angry' and 'kind' final prosody were duration (i.e., relatively shorter for anger and longer for kindness) and f<sub>0</sub> direction (i.e., a level for anger and a rise for kindness).

Since the 'angry' utterance was about 25% faster, and 10% higher in mean f<sub>0</sub> than the 'kind' utterance, the duration and f<sub>0</sub> initial point had to be slightly altered for the

sake of naturalness, when different style of the first part and final mora were combined. The duration was calculated using the original ratio of the final mora to the total utterance exclusive of pause. For example, the 'angry' duration for 'k' and 'a' was set to a value which was calculated by the total 'kind' utterance duration multiplied by the original 'angry' final mora ratio, when the final mora was combined with the 'kind' first part. The value of f0 initial point was set to the actual value of the style of the first part regardless of the style of the final mora. The values for these factors in each stimulus condition are summarised in Table 6.3. Although there was a slight difference in loudness between the 'angry' and 'kind' utterances, no amplitude adjustment was made. There were five occurrences for each stimulus condition. A total of 44 stimuli, consisting of five occurrences of eight conditions, and two dummy stimuli at the beginning and another two at the end, were prepared on a Sun workstation.

TABLE 6.3. Duration and f0 characteristics of the final mora ('ka') in each condition.

<i>CONDITION</i>			<i>Duration</i>		<i>F0</i>	
<i>First style*1</i>	<i>Final style*1</i>	<i>Final prosody</i>	<i>'k' 'a'</i>		<i>Initial point</i>	<i>Duration (rate of change in semitones/s)</i>
			<i>(in ms)</i>		<i>(Hz)</i>	
ANGRY	Angry	angry	50	60	160	Level (0)
	Angry	kind	60	80	160	Rise (70)
	Kind	angry	50	60	160	Level (0)
	Kind	kind	60	80	160	Rise (70)
KIND	Angry	angry	70	80	140	Level (0)
	Angry	kind	70	100	140	Rise (70)
	Kind	angry	70	80	140	Level (0)
	Kind	kind	70	100	140	Rise (70)

\*1: style differences were produced 'naturally', by instructing the speaker to vary style.



3. Rating scale

The bipolar 8-cm scales shown in Fig. 6.7 were used for anger, kindness, politeness and naturalness. Although the speaker was not instructed to speak them politely or impolitely, the politeness scale was included, because politeness is a key concept in any social interaction in Japanese society, and thus people are accustomed to make politeness judgement in any social situation. Anger and kindness are not generally considered on the same dimension as politeness. However, if we adopt such a definition of politeness as "a special way of treating people, saying and doing things in such a way as to take into account the other person's feelings" (Penelope Brown, 1980, p. 114), speaking in a kind/considerate way can be one way of being polite, and speaking in an angry/irritated way can be one realisation of impoliteness. The naturalness scale was also included to assess that all the stimuli (especially utterances which combined different speaking styles for the first part and the final mora) sounded reasonably natural (i.e., not too unnatural for rating any kind of affect studied).

NUMBER	CODE	SCALE		
		-VERY	NEUTRAL	+VERY
[   ]	[   ]	-----0-----		
	[   ]	-----0-----		
	[   ]	-----0-----		
	[   ]	-----0-----		

FIG. 6.7. Rating scales for anger, kindness, politeness and naturalness used in Experiment 2. CODE is the scale of affect: Anger, Kindness, Politeness or Naturalness.

#### *4. Subjects*

Nineteen paid subjects (11 male and 8 female), mostly postgraduates at British universities participated in the experiment. They were all native speakers of Japanese, ranging in age between 21 to 44 (average 28).

#### *5. Rating sessions*

Subjects were given written instructions about what they were supposed to do in the listening test. The instructions are in Appendix G. Each utterance was presented to subjects in random order over a small loudspeaker attached to a Sun workstation under their own control by using a keyboard key. They were asked to rate the utterance on all four scales of politeness, anger, kindness and naturalness, starting with the scale which they felt most appropriate for the utterance they had just heard. They had a practice session with four stimuli consisting of the speaker's 'angry', 'kind' and 'neutral' utterances and the 20% compressed version of the 'kind' utterance. All the subjects were tested individually in a quiet room and each session lasted about 20 minutes.

#### **6.1.2.2. Results and discussion**

The scores were obtained by measuring the distance between subjects markings on the linear scale and a mid point on the scale. Scores could range between -4 (- very) and +4 (+ very). Following a keyboard response, the time was recorded by a computer program which controlled stimulus presentation, and the difference between the successive times in milliseconds was used as a rough indication of reaction time. The mean values and standard deviations over 19 subjects' five repetition scores for politeness, anger, kindness and reaction time are shown in Table 6.4.



TABLE 6.4. Mean values and SDs over 19 subjects' five repetition scores for politeness, anger, kindness and reaction time.

<i>CONDITION</i>			<i>Polite</i>	<i>Angry</i>	<i>Kind</i>	<i>Natural</i>	<i>Reaction Time</i>
<i>First style</i>	<i>Final style</i>	<i>Final prosody</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>
ANGRY	Angry	angry	-1.28 (0.93)	1.73 (1.09)	-2.09 (0.89)	0.98 (1.61)	18.8 (3.2)
		kind	-0.61 (0.74)	1.04 (0.94)	-1.28 (1.00)	1.07 (1.19)	20.8 (5.5)
	Kind	angry	0.21 (0.77)	-0.47 (1.27)	-0.68 (0.90)	1.38 (1.18)	19.6 (4.7)
		kind	0.67 (1.35)	-0.99 (1.62)	-0.04 (1.20)	1.45 (1.48)	19.6 (5.2)
	Angry	angry	-0.27 (1.36)	0.91 (1.06)	-1.63 (0.93)	-0.33 (1.79)	20.3 (4.9)
		kind	-0.44 (1.36)	1.10 (1.07)	-1.73 (1.03)	-1.01 (1.86)	19.4 (4.5)
KIND	Kind	angry	1.73 (0.98)	-2.14 (1.41)	0.97 (1.44)	0.74 (1.73)	17.7 (3.1)
		kind	2.14 (1.06)	-2.43 (1.06)	1.69 (1.09)	0.26 (1.73)	17.1 (4.1)

As was expected, there was a very high positive correlation between politeness and kindness (the Pearson  $r = 0.98$ ,  $p < 0.0005$ , 1-tailed), and a very high negative correlation between politeness and anger ( $r = -0.98$ ,  $p < 0.0005$ , 1-tailed). There was virtually no correlation between politeness and naturalness ( $r = 0.07$ ). Discussions on the anger/kindness rating scores would not give any additional information due to the very high level of correlation between politeness and anger/kindness. Therefore, only ratings of politeness and naturalness, together with reaction time, will be discussed in detail.

Kendall's coefficient of concordance ( $W$ ) was calculated to assess the inter-judge agreement among 19 judges' ratings of the eight conditions. There was a high level of agreement for politeness ( $W = 0.74$ ; the mean Spearman rank-order correlation coefficient ( $\text{Mean } r_s$ ) = 0.73,  $N = 19$ ), and moderate level of agreement for naturalness ( $W = 0.39$ ;  $\text{Mean } r_s = 0.36$ ,  $N = 19$ ) ( $p_s < 0.001$ ). The effective reliability ( $R$ ) was also very high for both politeness and anger ( $R > 0.9$ ). Since the mean reliability was not very high for naturalness, the intra-judge agreement was assessed by the ratio between variance of each judge's repetition scores to the total variance (Section 3.5.2.5). An ANOVA test was performed on naturalness scores with factors of speaking style of the first part (two styles), speaking styles of the final mora (two styles), prosody of the final mora (two types) and each subject's repetition factor (five repetitions). It was found that the repetition factor was non-significant. This result shows that each subject's ratings were consistent. However, this consistency may not ensure that the rating scale method works in distinguishing fine levels of naturalness, because some subjects' scores were all 0 (neutral) or 4 (very natural).

ANOVA tests were performed on the mean values of five scores for politeness, naturalness and reaction time for each condition separately. The within-subject factors were two speaking styles of the first part (First Style), two speaking styles of the final mora (Final Style) and two types of prosody of the final mora (Final Prosody). To assess the relative magnitude of the effects of each factor, the eta-squared was calculated. All results of these ANOVA tests are attached in Appendix 2 (2), and the significant effects at the level of 0.05 or better are shown in Table 6.5.



TABLE 6.5. ANOVA results of Experiment 2: significant effects at the level of 0.05 or better.

(a) Politeness

	<i>F</i>	<i>p: significance of F</i>	<i>eta-squared (%)</i>
<u>MAIN EFFECTS</u>			
First Style	40.13	**	14.1
Final Style	45.03	**	43.6
Final Prosody	5.32		1.5
<u>INTERACTIONS</u>			
First Style and Final Style	23.97	**	2.7
First Style and Final Prosody	5.87		0.7
First Style and Final Style and Final Prosody	5.00		0.5

$df_{effect} = 1, df_{error} = 18$

\*\* :  $p < 0.001$

TABLE 6.5. (continued)  
(b) Naturalness

	<i>F</i>	<i>p: significance of F</i>	<i>eta-squared (%)</i>
<u>MAIN EFFECTS</u>			
First Style	16.62	**	23.3
Final Style	11.90	*	8.3
<u>INTERACTION</u>			
First Style and Final Prosody	6.53		1.5

$df_{effect} = 1, df_{error} = 18$   
\* :  $p < 0.01$   
\*\* :  $p < 0.001$

(c) Reaction time

	<i>F</i>	<i>p: significance of F</i>	<i>eta-squared (%)</i>
<u>MAIN EFFECTS</u>			
First Style	5.82		4.8
Final Style	12.22	*	7.1
<u>INTERACTIONS</u>			
First Style and Final Style	4.55		5.0
First Style and Final Prosody	5.80		3.2

$df_{effect} = 1, df_{error} = 18$   
\* :  $p < 0.01$



### (a) Politeness

The analysis showed significant main effects of First Style, Final Style and Final Prosody, and significant interactions between First Style and Final Style, and First Style and Final Prosody, and the significant three-way interaction of the main factors ( $ps < 0.05$ ) (Table 6.5 (a)). Among them, the effect of Final Style was found to be the most influential and Final Prosody was less important, according to the eta-squared.

Fig. 6.8 clearly shows the importance of the speaking style of the final mora.

Utterances with the 'angry' Final Style were rated negatively regardless of the style of the rest of the utterance. The speaking style of the first part contributed to the judgements of the politeness degree. In other words, if the 'kind' Final Prosody was combined with the 'kind' First Style, the utterance was judged as more polite than the utterance of the 'kind' Final Prosody with the 'angry' First Style.

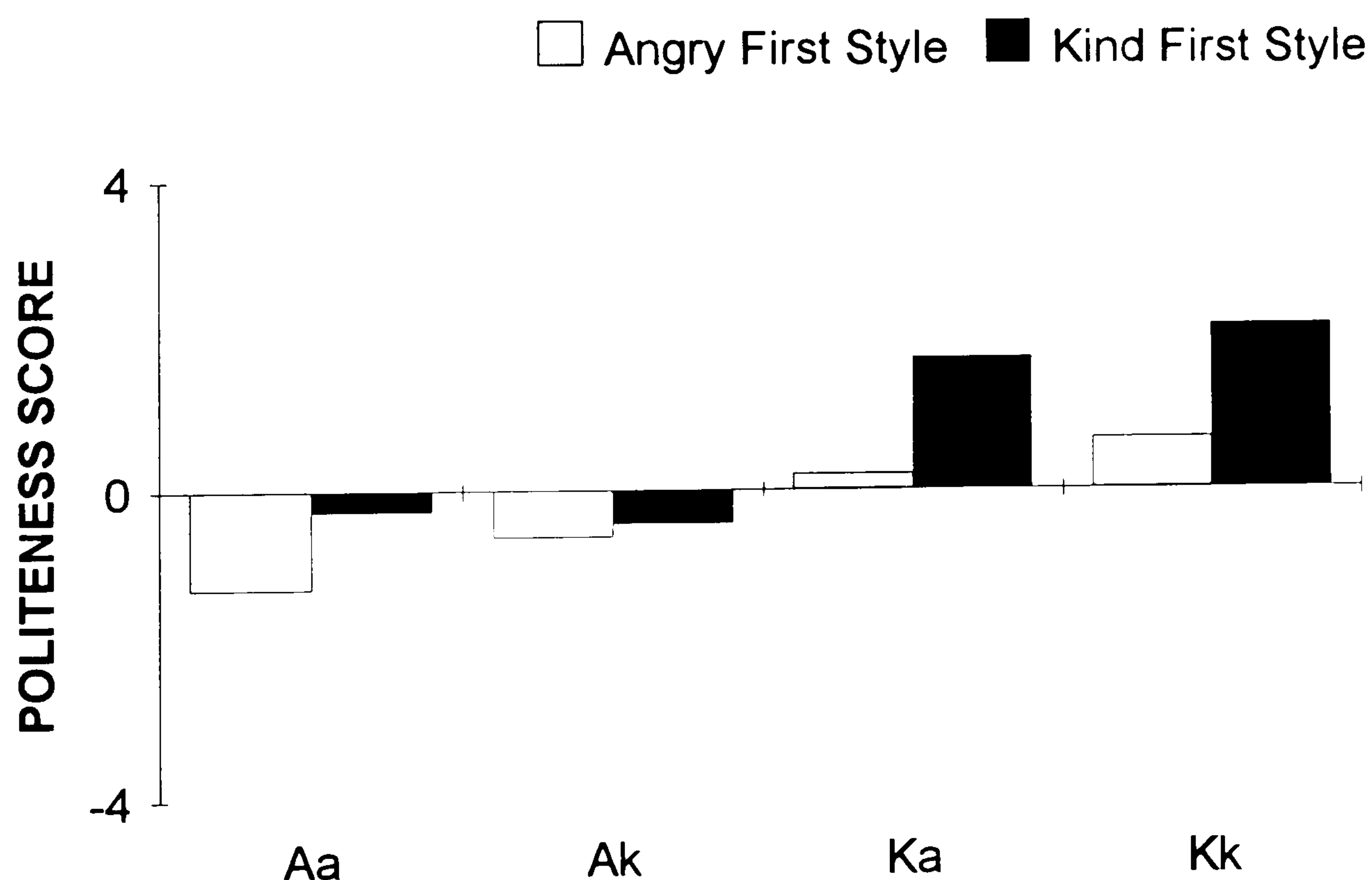


FIG. 6.8. Mean ratings for politeness by speaking style of the first part (First Style), speaking style of the final mora (Final Style) and prosody of the final mora (Final Prosody) in Experiment 2. 'Aa' means 'Angry' Final Style with 'angry' Final Prosody; 'Ak', 'Angry' Final Style with 'kind' Final Prosody; 'Ka', 'Kind' Final Style with 'angry' Final Prosody; 'Kk', 'Kind' Final Style with 'kind' Final Prosody.

### (b) Naturalness

There were significant main effects of the speaking style of both the first part and final mora, and the significant interaction between First Style and Final Prosody ( $ps < 0.05$ ) (Table 6.5 (b)). The eta-squared showed that the effect of the style of the first part was the most salient.

The mean naturalness scores in relation to the mean politeness scores for each condition are shown in Fig. 6.9. The figure shows that the utterances with the 'kind' first style were rated as less natural than the 'angry' counterparts. There was no significant difference in naturalness between the four conditions of the 'angry' first style. Among the eight conditions, the 'kind' first style combined with the 'angry' final style was rated least natural, as was expected. This may be because the 'kind' source utterance sounded slightly exaggerated (i.e., sounding as if the speaker were talking to a senior citizen or a small child), and there was a gap in loudness between the 'kind' final mora and the 'angry' final mora (i.e., the 'angry' final mora was spoken much more loudly than the 'kind' final mora). However, the other conditions which also combined different speaking styles (i.e., the 'angry' first style combined with the 'kind' final style) were rated most natural among the eight conditions. Although we have seen that there was no correlation between the natural scores and politeness scores of the eight conditions ( $r = 0.07$ ), when each first style (i.e., either angry or kind) is examined separately, there appears to be a tendency for less natural stimuli to sound less polite. (Fig. 6.9). The role of naturalness in politeness judgements will be discussed in more detail in Experiment 3 (Section 6.3).



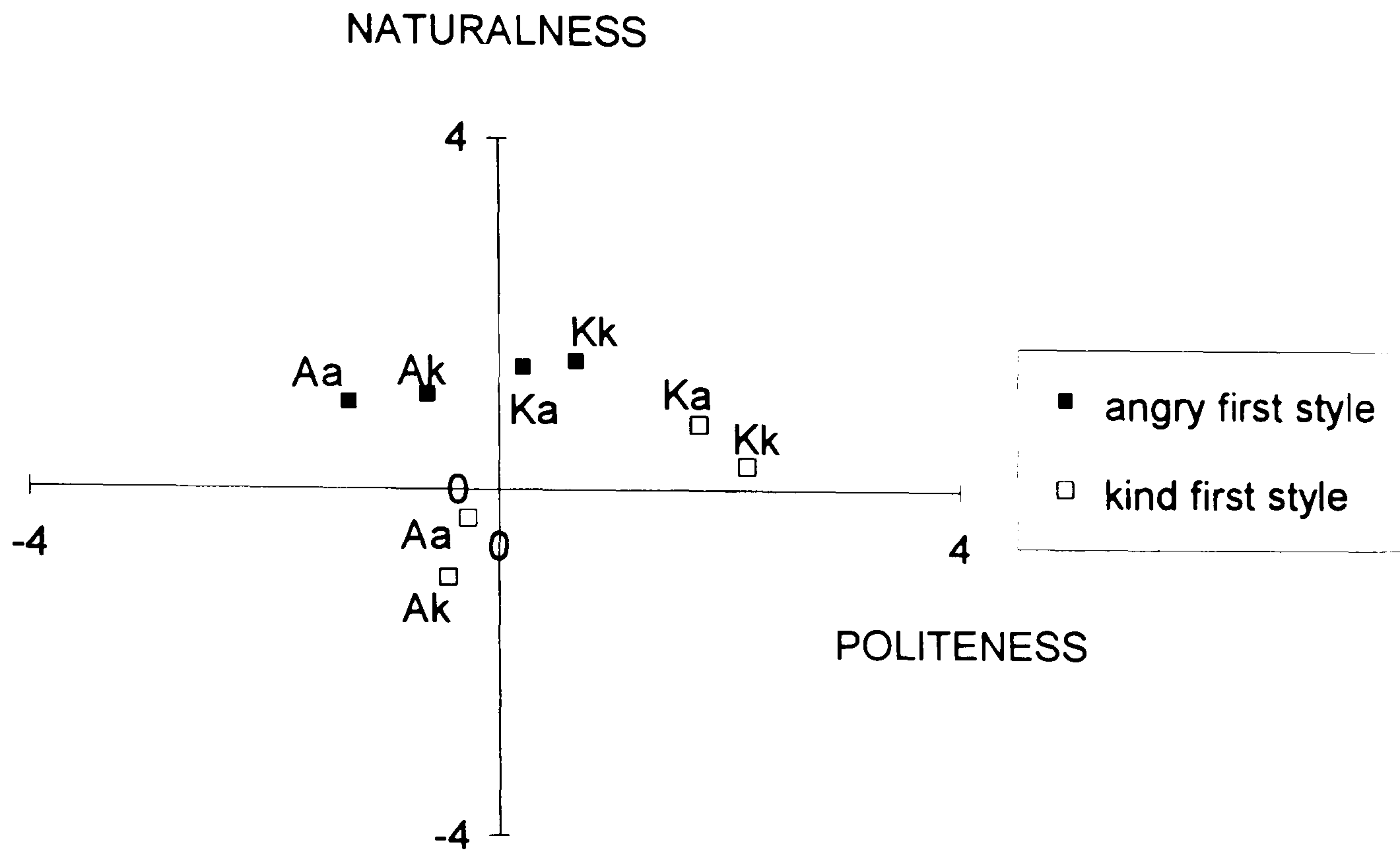


FIG. 6.9. Mean politeness and naturalness scores for each condition in Experiment 2. The 'angry' and 'kind' first styles are shown separately. 'Aa' means 'Angry' final style with 'angry' final prosody; 'Ak', 'Angry' final style with 'kind' final prosody; 'Ka', 'Kind' final style with 'angry' final prosody; 'Kk', 'Kind' final style with 'kind' final prosody.

### (c) Reaction time

The ANOVA results on the scores of reaction time showed significant main effects of the speaking style of both the first part and final mora, and significant interactions between these styles, and the style of the first part and the final prosody ( $ps < 0.05$ ) (Table 6.5 (c)). Since the difference between the style of the first part and the style of the final mora was noticeable, it was expected that the inconsistency between the style of the first part and the style of the final mora would be the most salient factor for reaction time. This, however, was not supported: the most salient factor was the style of the final mora. Subjects appear to have primarily responded to the style of the final mora.

### **6.1.3. Potential effects of f0 manipulation on voice quality**

Experiment 2, which was reported in the previous section, examined the effects of speaking style which includes vowel and voice quality, by using computer cue manipulated stimuli. However, computer cue manipulation often causes spectral changes in some degree, and thus could change vowel and voice quality to such an extent that it introduces undesirable artefacts into the manipulated stimuli. Therefore, auditory and spectral analyses were carried out to assess potential effects of f0 manipulation by the TD-PSOLA technique on voice quality.

#### **6.1.3.1. Vowel production**

Long vowels, 'ae' and 'ah', were produced by a male speaker, who was well experienced in controlling his voice being a phonetician himself, with five different tones: level, rise, fall, rise-fall and fall-rise. The vowels with a level tone and the vowels with a rising tone ('human-rise' versions) were used for spectral analysis. They were recorded on a SONY TCD-D7 DAT recorder, and then digitised at a sampling rate of 16 kHz onto a Sun Workstation via its own A-to-D boards. The schematic figures of f0 movements of these vowels with a level tone and a rising tone are shown in Fig. 6.10. The computer manipulated versions (PSOLA-rise versions) were produced by modifying f0 and duration of these level tone versions by a computer program based on the TD-PSOLA algorithms. F0 and duration values were manipulated in such a way that they realised the f0 movement of the human-rise versions by linear interpolation with two lines.



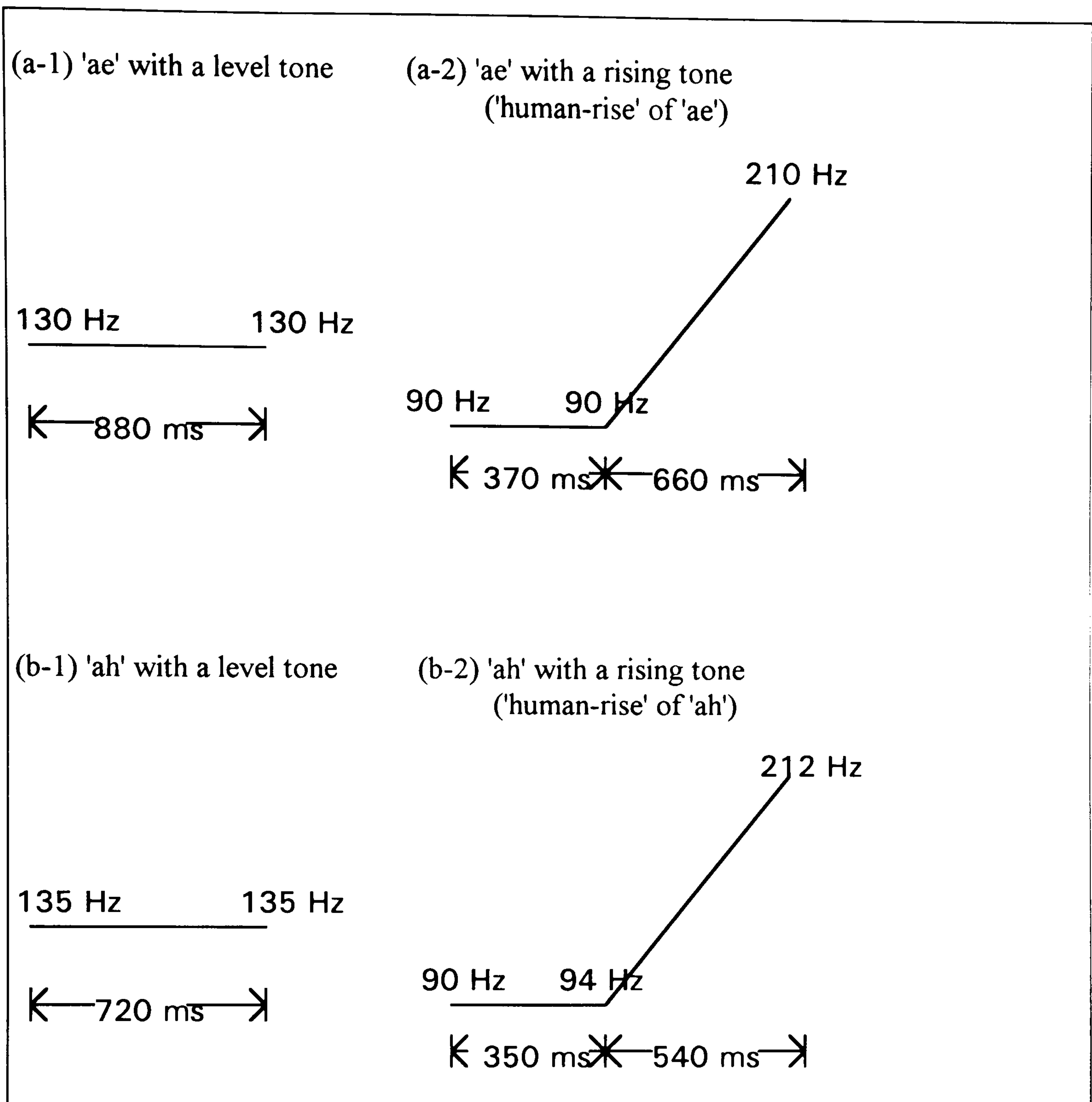


FIG. 6.10. Schematic figures of  $f_0$  movements of two vowels with a level tone and a rising tone spoken by one male speaker. The  $f_0$  movement of vowel 'ae' (a) and vowel 'ah' (b) are shown separately.

### 6.1.3.2. Spectral analysis and auditory evaluation

The formant structures of the vowels with a level tone, the human-rise vowels and the PSOLA-rise vowels were analysed in terms of formant frequency, bandwidth and intensity. Spectral analyses were performed as follows. Spectral slices were taken from the middle of the segment for the level vowels, and from one point whose  $f_0$  was 130 Hz and from the other point whose  $f_0$  reached 190 Hz for the rise vowels. A smoothened spectral envelope was achieved using an LP analysis technique with an autocorrelation method of the digital signal processing software package ESPS/Waves (Entropic Research Laboratory, 1993). Estimated formant frequencies and bandwidths for these vowels are shown in Fig. 6.11.

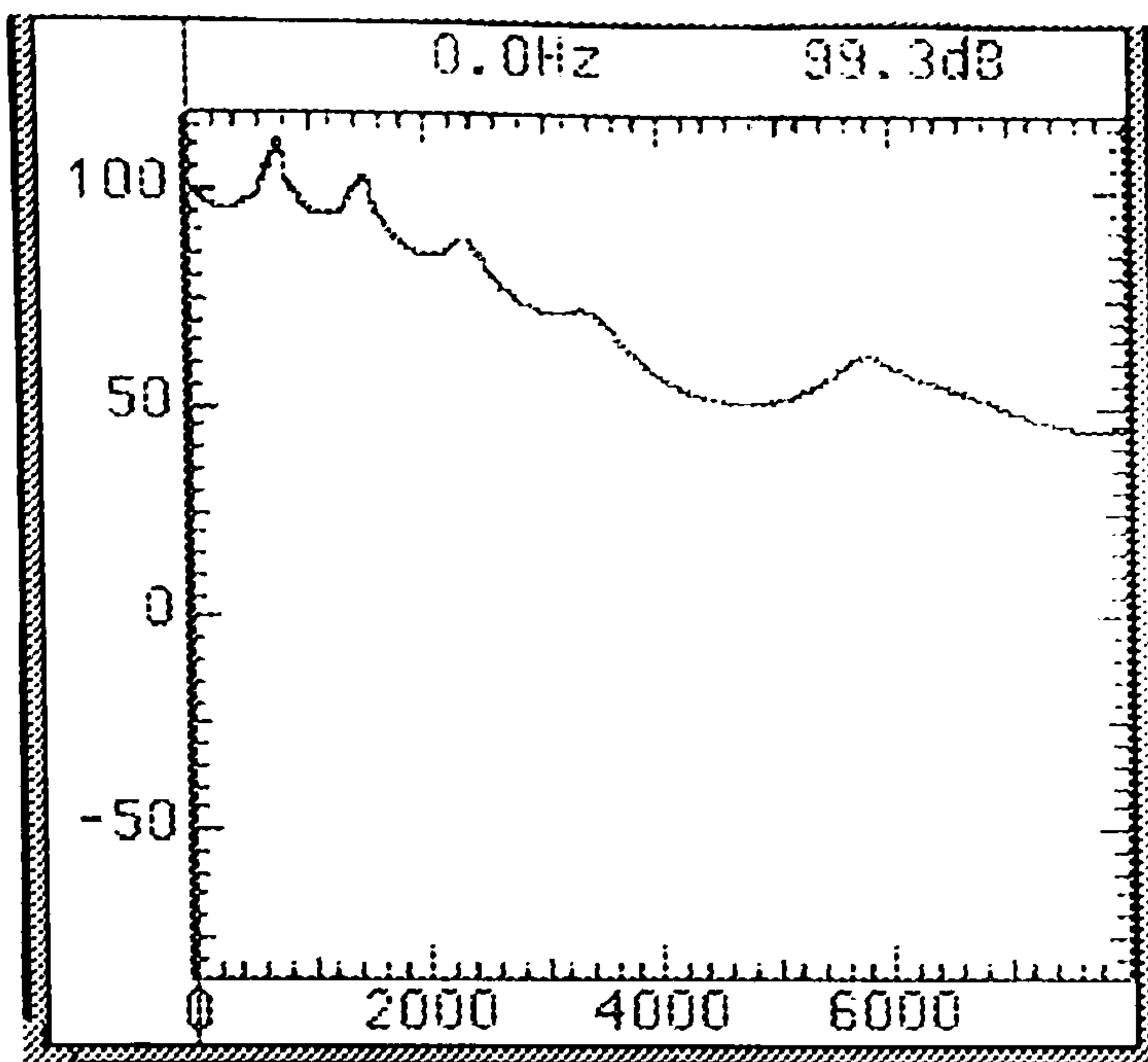
The formant structures show that changes of  $f_0$  appear to have effects, which are slight but not fully predictable, on upper formants. This is true of both the human speaker and of PSOLA. There has to be a little caution for potential effects caused by manipulation in terms of voice quality, in view of slight changes of upper formants.

However, these seem to be rather minor in terms of perception of voice and vowel quality. The auditory assessment confirmed that the PSOLA manipulated vowels retained the voice quality of the speaker very well. In fact, when the original level vowels were compared with the human-rise vowels and with the PSOLA-rise vowels, the PSOLA-rise vowels, especially for the vowel 'ae', sounded nearly the same as the original level ones, while the human speaker did change a voice quality to some extent. In both manipulations (human and PSOLA), the vowel qualities were very well preserved.

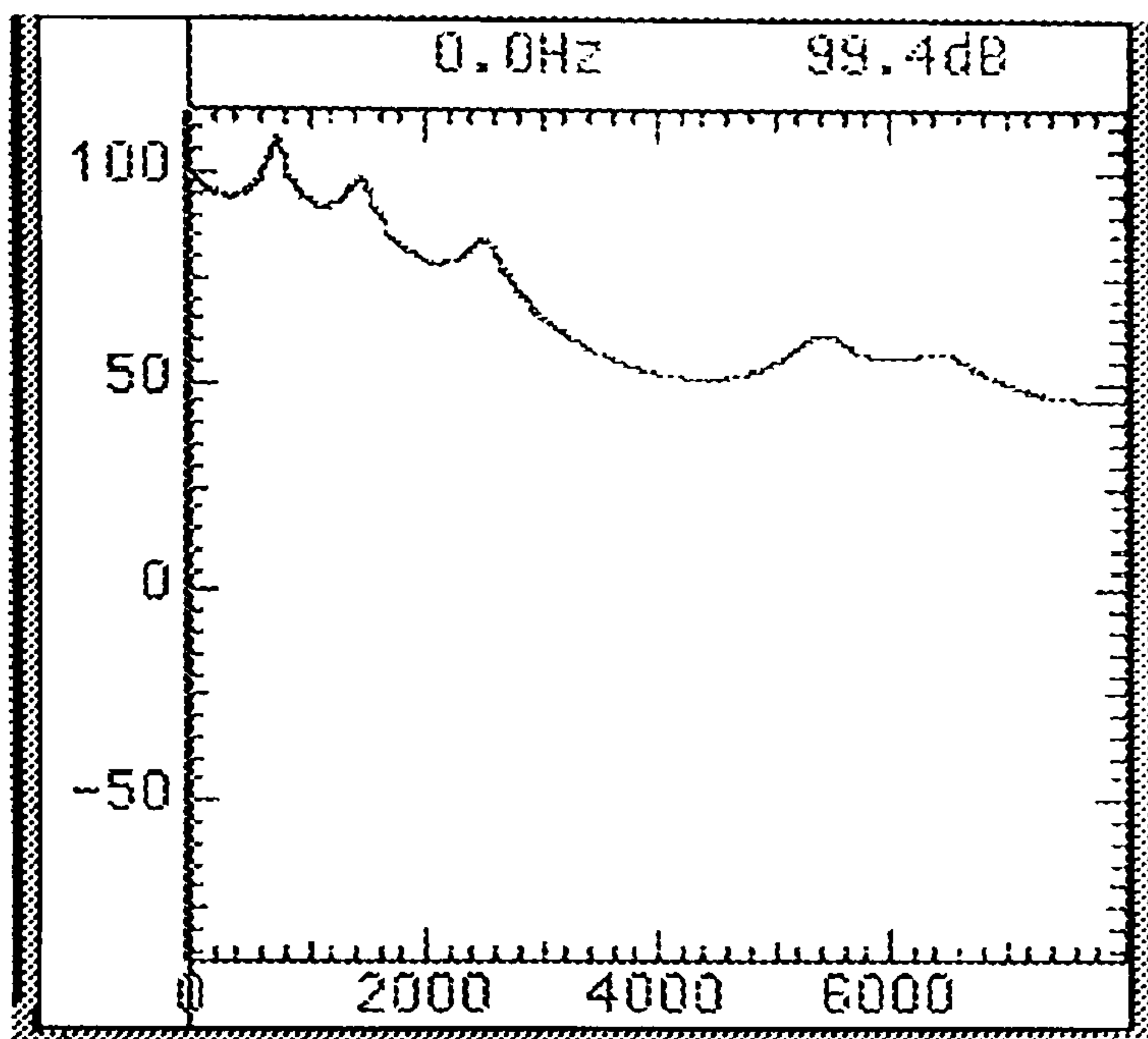


(a) VOWEL: 'ae'

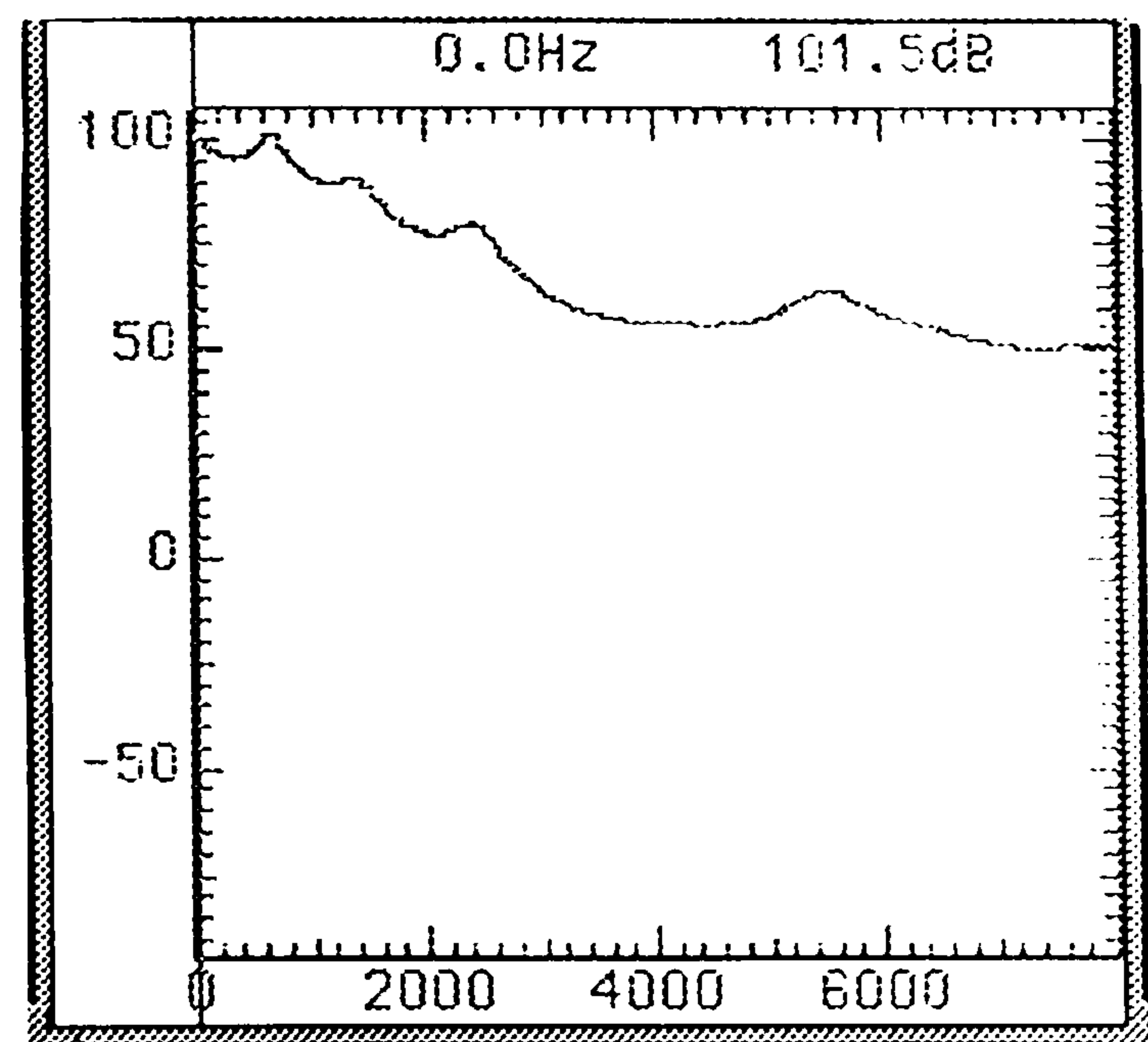
(a-1) Level tone (at around the middle of the vowel,  $f_0 = 130$  Hz)



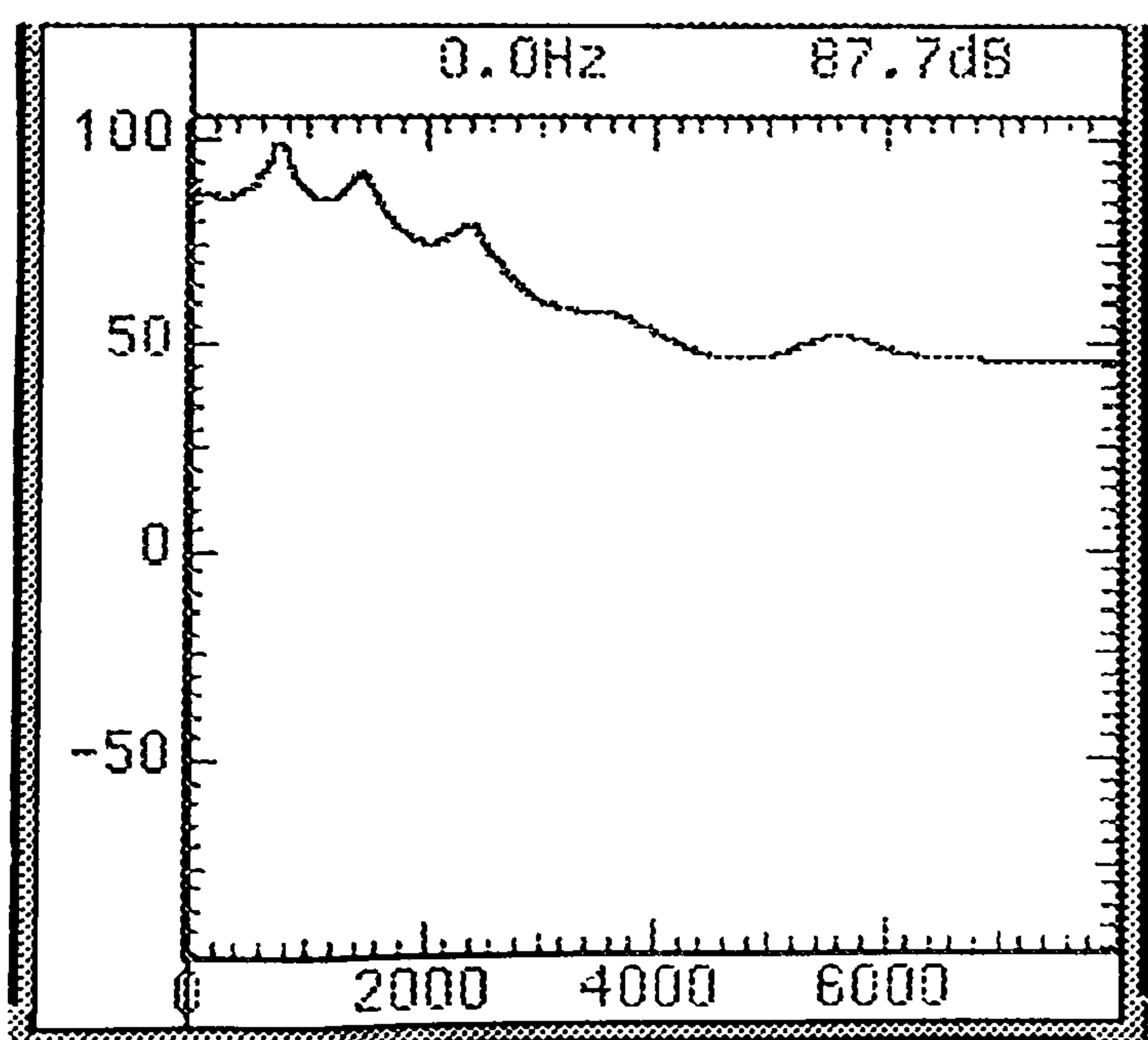
(a-2) HUMAN-RISE ( $f_0 = 130$  Hz)



(a-3) HUMAN-RISE ( $f_0 = 190$  Hz)



(a-4) PSOLA-RISE ( $f_0 = 130$  Hz)



(a-5) PSOLA-RISE ( $f_0 = 190$  Hz)

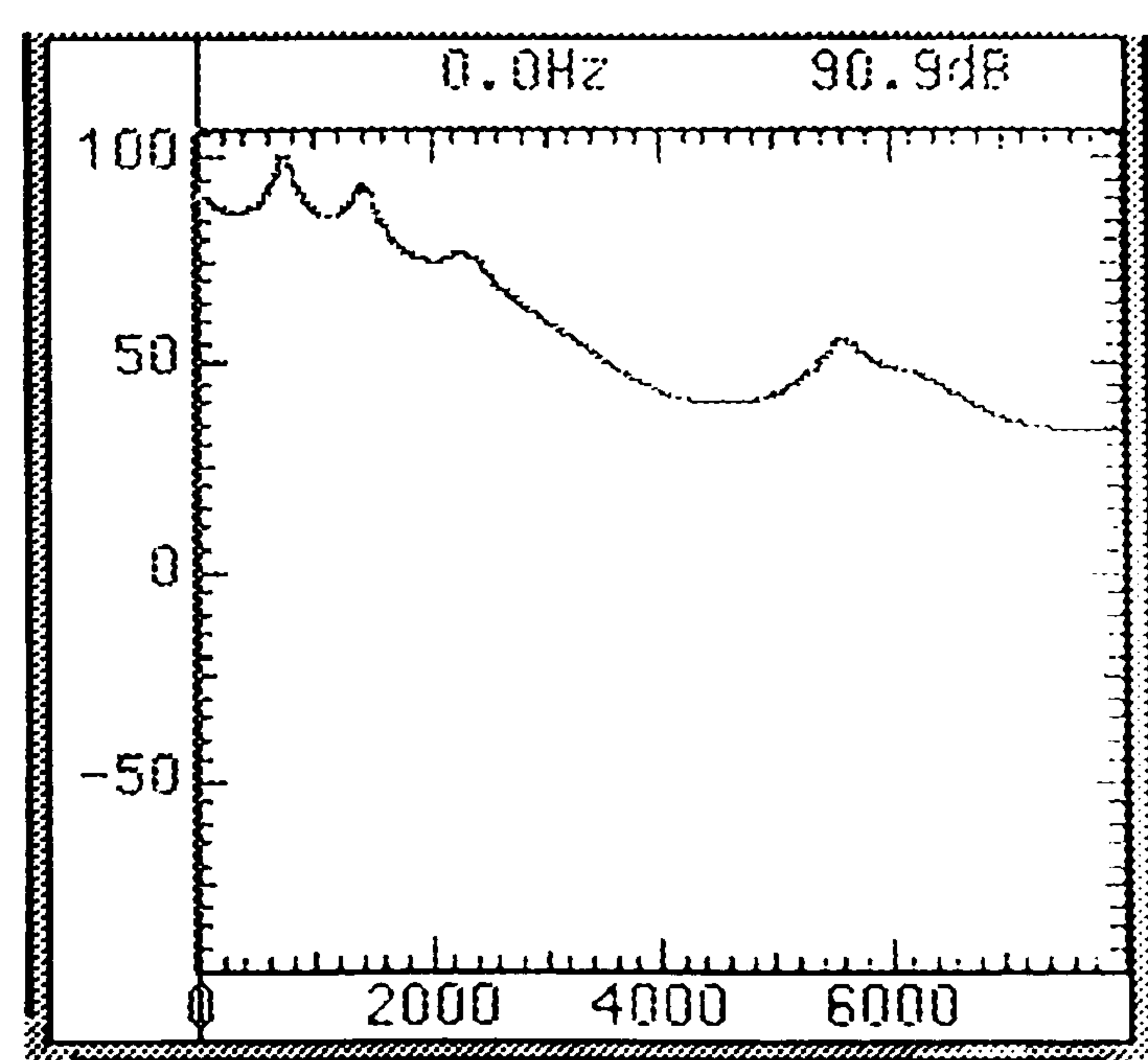
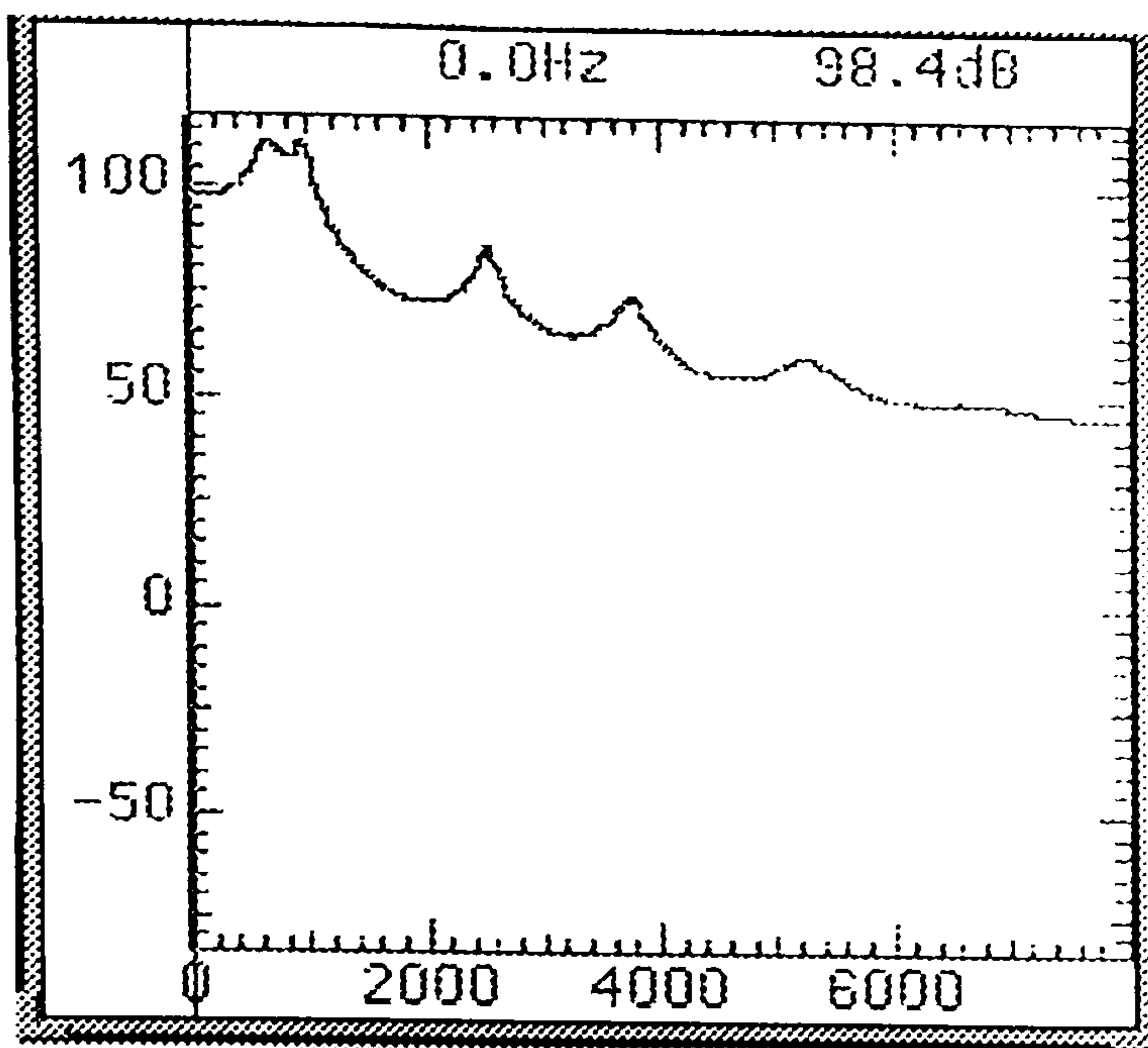


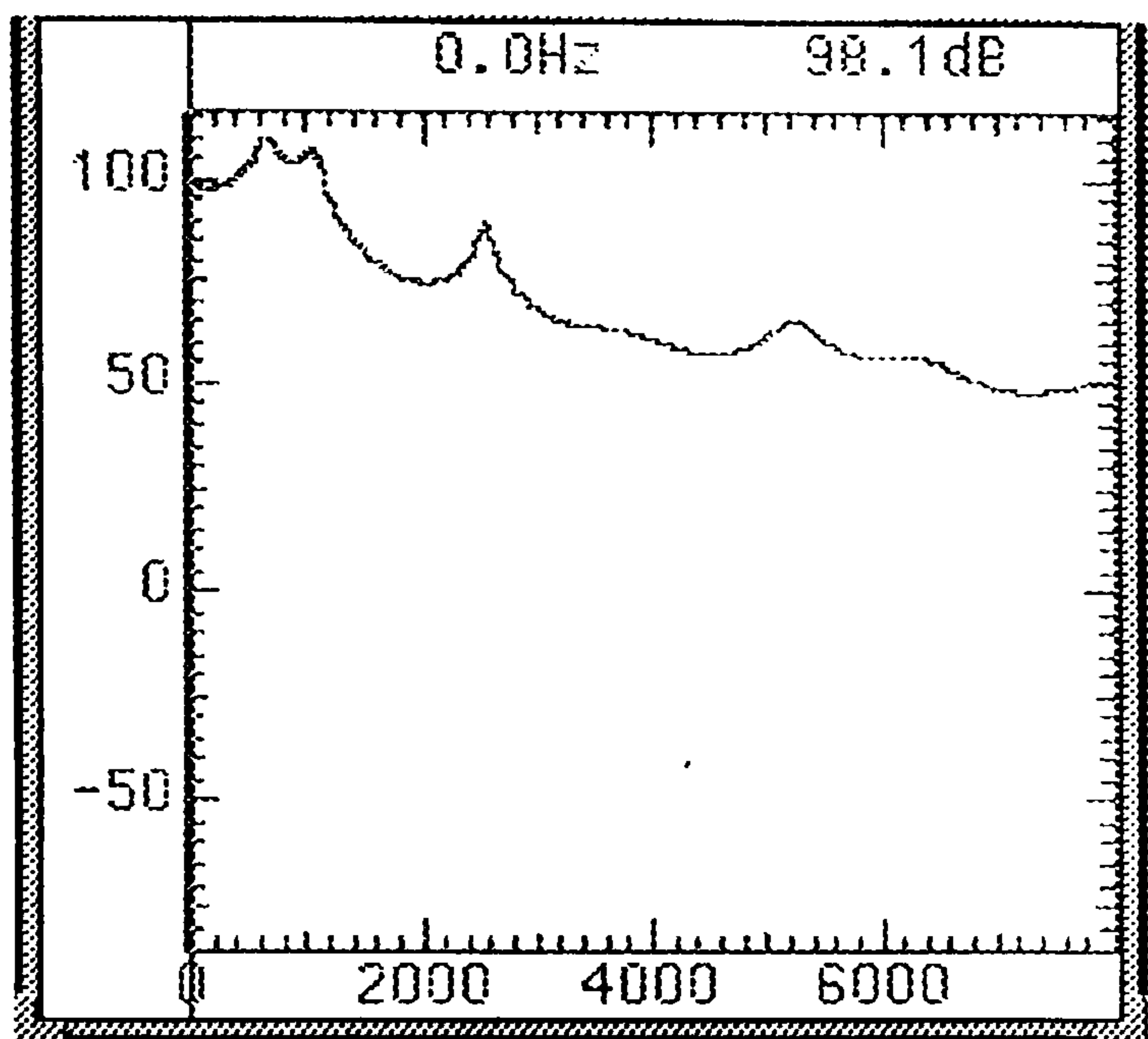
FIG. 6.11. Estimated formant structures for two vowels: 'ae' (a) and 'ah' (b). The x-axis is frequency in Hz, and the y-axis is amplitude in dB.

(b) VOWEL: 'ah'

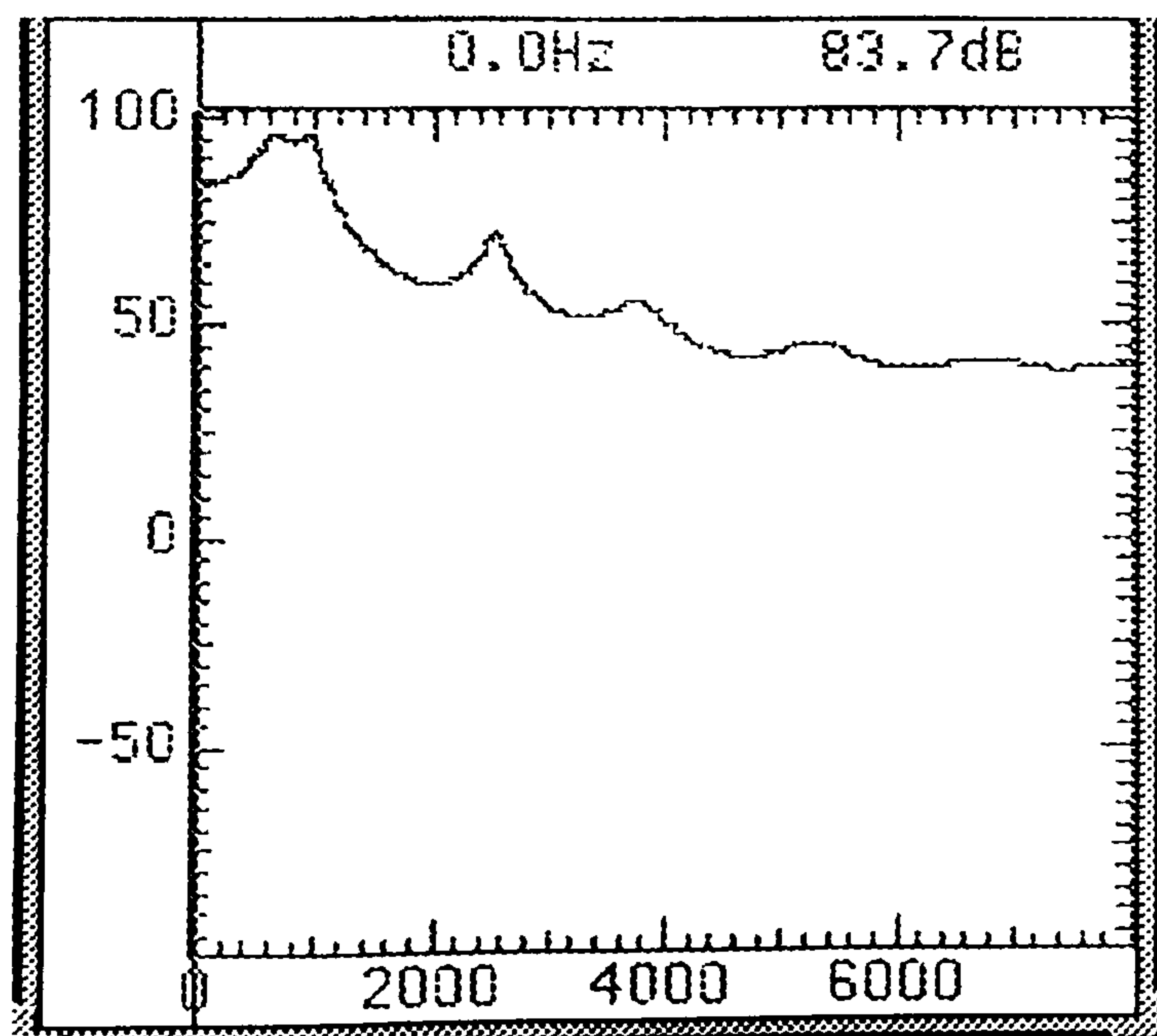
(b-1) Level tone (at around the middle of the vowel,  $f_0 = 130$  Hz)



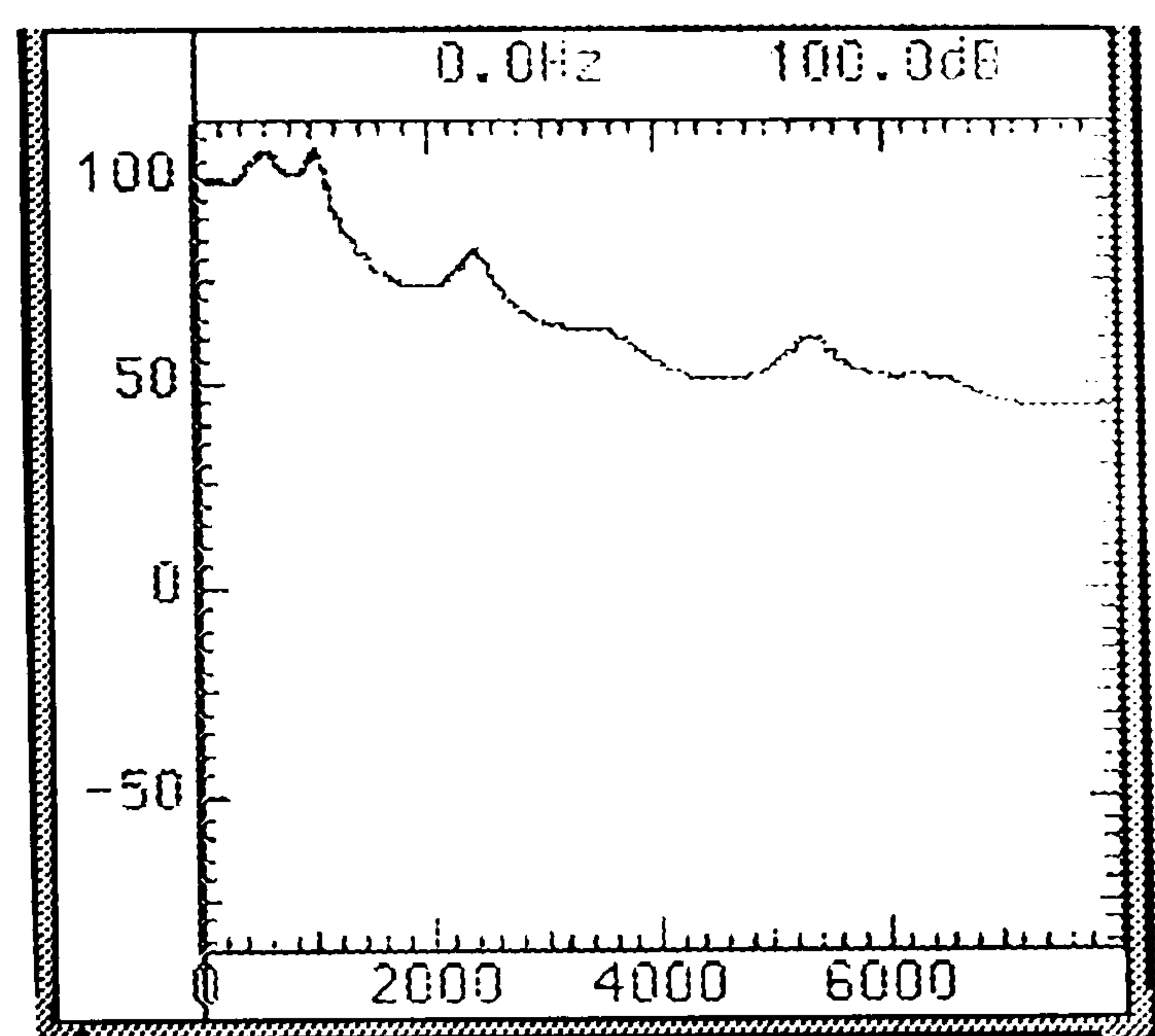
(b-2) HUMAN-RISE ( $f_0 = 130$  Hz)



(b-4) PSOLA-RISE ( $f_0 = 130$  Hz)



(b-3) HUMAN-RISE ( $f_0 = 190$  Hz)



(b-5) PSOLA-RISE ( $f_0 = 190$  Hz)

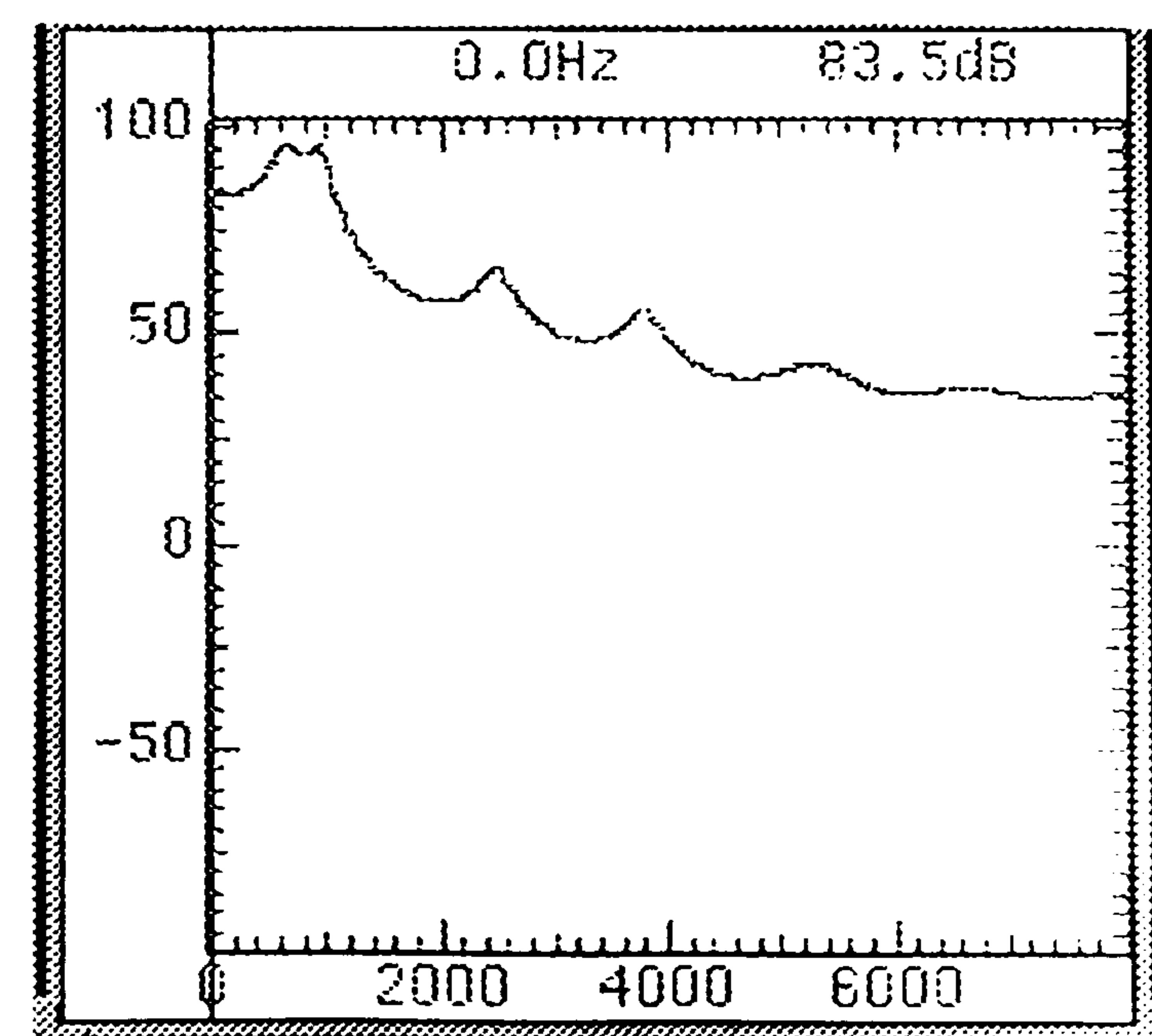


FIG. 6.11. (continued)



#### 6.1.4. Summary of the experiments on the final part

Both Experiments 1-A and 2 confirmed the importance of the final part of the utterance in signalling affect. The effects of the final mora were so great that how the final mora ('ka') was spoken determined the overall impression of the utterance regardless of the speaking style of the rest of the utterance. In Experiment 1-A prosody of the final mora was found to be very influential in ratings of politeness. Relatively shorter durations (the ratio of the final mora to the total utterance duration exclusive of pause: 1.7) with a rising tone were rated more positively than longer final durations (the final mora ratio: 2.2) with a falling tone. In Experiment 2 three factors were investigated: speaking style of the final mora (i.e., how the final mora was spoken, except  $f_0$  movement), prosody of the final mora (i.e.,  $f_0$  movement) and speaking style of the rest of the utterance. Relatively shorter durations (the final mora ratio: 1.2) with a level tone were adopted as the 'angry' (non-polite) final prosody, and relatively longer final durations (the final mora ratio: 1.5) were as the 'kind' (polite) final prosody. Among these factors, the speaking style of the final mora was found to be the most important, and the final prosody was the least important.

Experiment 2 was concerned with speaking style, which is a complex of features including articulation and voice quality, and it has been known that computer prosodic cue manipulation involves spectral changes to some degree. Therefore, auditory and spectral analyses were carried out to examine potential effects of  $f_0$  manipulation by the PSOLA technique, on vowel and voice quality, using two long vowels (Section 6.1.3). The spectral analyses showed that there were slight changes in the upper formant structures, thus a slight change in voice quality. However, these effects were minor in terms of auditory impressions. The structures of the first three formants which determine vowel quality were very well preserved.

In summary, it can be concluded that the final part of the utterance has a great

effect on human perception of affect. However, it is not clear whether or not prosody (or f0 movement) is less important than speaking style, because it appears that people react to the most salient factor. The difference in final duration was most noticeable in Experiment 1-A while the style of the final mora was the most salient feature in Experiment 2, mainly due to the difference in massiveness of manipulation.

## **6.2. Experiment 1-B: The role of speech rate**

Speech rate has been studied in relation to various kinds of attitudinal meanings including politeness. Brown *et al.* (1974) used 15 paired opposite adjectives, which were later clustered into three factors based on the patterns of the scores, as rating scales: 'benevolence' including polite, kind, sincere, religious and just; 'competence' including active, ambitious, intelligent and confident; and 'others' including happy, good-looking, strong, dependable, sociable and likeable. A series of studies found that rate manipulation had much greater and consistent effects than f0 mean and variation manipulations, especially for competence factors, but less clear effects for benevolence (Brown *et al.*, 1974; Smith *et al.*, 1975). The major finding about the effects of speech rate is that increased rate increased competence and decreased benevolence (Brown *et al.*, 1974; Smith *et al.* 1975). Studies with American subjects showed that benevolence had an inverted-U shape as a function of rate (i.e., having the maximum scores at the normal rate) (Brown *et al.*, 1974; Smith *et al.*, 1975; Bruce Brown, 1980). However, another study with British subjects found that decreased rate linearly increased benevolence (Brown, Giles and Thakerar, 1985).

The findings mentioned above suggest that (1) speech rate can influence politeness ratings substantially and (2) slower or normal speech rate contribute to higher politeness scores. Experiment 1-B was conducted to investigate these points. There is one factor which needs careful consideration for designing any experiments



on speech rate: the selection of the range of speech rate of stimulus utterances. Speech rates which are abnormally slow or fast are not realistic, and therefore it cannot be expected to obtain insightful findings with such unrealistic stimuli. In order to assess the normality of the speech rate of the stimuli, a listening test was carried out prior to Experiment 1-B. The listening test is reported in Section 6.2.1 and Experiment 1-B is reported in Section 6.2.2.

### 6.2.1. Pre-test: Listening test for assessing the normality of speech rates

#### 6.2.1.1. Procedure

The extremity of the speech rate will certainly influence politeness ratings. In order to assess the normality of the range used in the main experiment, an informal listening test was conducted with three Japanese subjects, including the author, with the rating scale shown in Fig. 6.12.

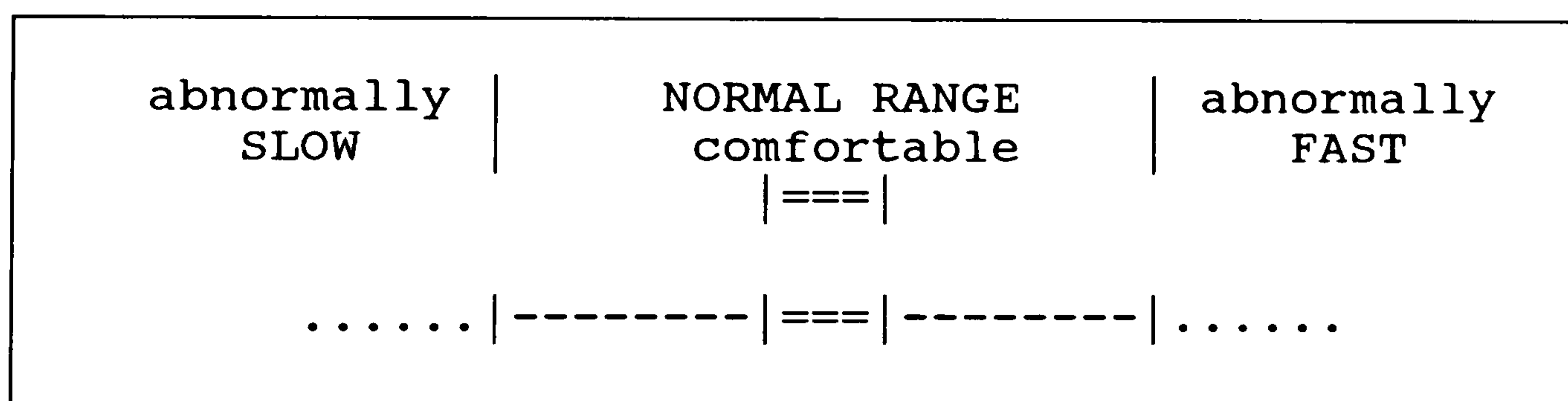


FIG. 6.12. Scale used in the pre-test.

The stimuli used were the polite and casual versions of the 'luggage' sentence spoken by three speakers, including the two speakers (KS and TK), who were used in Experiment 1-A (Section 6.1.1) and Experiment 1-B (which is reported in the next section). Changes in segmental duration rates ranged from 0.7 to 1.3. The actual speech rates for the three speakers were 10.8, 11.8 and 11.8 mora per sec. The third speaker, SF, whose speech rate was the same as that of TK's utterance,

was included to examine whether or not perceived tempo would substantially vary from speaker to speaker. The rate manipulation was performed using a computer program based on the PSOLA technique. Each stimulus was presented through headphones only once in random order, preceded by a beep tone and followed by a 4-second silence during which subjects were asked to rate the utterance on the scale of tempo.

#### 6.2.1.2. Results

Perceived tempo index was calculated as a normalised position of the markings on the scale. The normalisation was carried out in such a way that 'extremely slow (but in an acceptable range)' was set to 0, and 'extremely fast' was set to 100. Negative scores mean 'abnormally slow' and scores over 100 mean 'abnormally fast'. The scores between 40 and 60 are in the 'comfortable range' of the raters. The perceived tempo index (PTX) and speech rate exclusive of pause and final morae in Phrase 1 and Phrase 2 (SR) in mora per sec of the polite versions are shown in Table 6.6.

The findings are as follows. Firstly, the intercorrelations ( $r$ ) between speech rate (in mora per sec) and the perceived tempo index of all the polite utterances of the three speakers were very high for all three subjects ( $r = 0.9$ , 2-tailed test,  $ps < 0.001$ ,  $N=27$ ). Secondly, the perceived tempo (and therefore the comfortable range) varied from speaker to speaker. For KS, his actual rate to the 20% compressed version (Speech rate: 10.8 ~ 13.5) were rated as comfortable by at least two subjects out of three. The 5% to 10% compressed versions (Speech rate: 12.3 ~ 13.0) were perceived as comfortable for TK, while the 10% expanded version to his actual rate (Speech rate: 11.2 ~ 11.8) were comfortable for SF. An interesting thing is that although TK and SF had the same speech rate, the faster versions (F05 ~ F10) were preferred for TK whereas the slower versions (S10 ~ Unmodified rate)



were preferred for SF. This may be due to the difference in their articulation: TK tended to devoice or reduce vowels while SF spoke vowels very clearly. This articulation difference may have caused differences in perceived rhythm and also the degree of degradation caused by computer manipulation. Finally, none of the utterances in the change rates between 0.8 and 1.2 were rated as 'abnormal', except one utterance: SF's utterance with the rate of 0.8 (F20) was rated as 'abnormally fast' by one subject.

TABLE 6.6. Perceived tempo index (PTX) and speech rate (SR) for the nine speech rate levels (S30 ~ F30) of the polite versions of the 'luggage' sentence spoken by three male speakers (KS, TK and SF). The scores by three subjects (Sub 1 ~ Sub 3) are shown separately. 'Abnormally slow/fast' are highlighted by shading.

<i>speaker</i>		<i>S30</i>	<i>S20</i>	<i>S10</i>	<i>S05</i>	<i>UM</i>	<i>F05</i>	<i>F10</i>	<i>F20</i>	<i>F30</i>
KS	<i>SR</i>	8.3	8.9	9.8	10.2	10.8	11.4	11.9	13.5	15.4
	Sub1	14	29	43*	43*	46*	54*	64+	57*	75
	Sub2	-11	14	20	18	48*	50*	48*	89	80
	Sub3	-7	20	30	27	29	50*	50*	57*	84
TK	<i>SR</i>	9.0	9.8	10.6	11.2	11.8	12.3	13.0	14.7	16.7
	Sub1	23	34	34	48*	50*	50*	48*	71	80
	Sub2	12	12	32	32	27	48*	48*	48*	104
	Sub3	2	4	30	34	23	48*	62+	68	77
SF	<i>SR</i>	9.0	9.8	10.6	11.2	11.8	12.3	13.0	14.7	16.7
	Sub1	30	43*	41*	50*	50*	54*	73	73	93
	Sub2	14	18	27	48*	48*	95	91	111	112
	Sub3	23	18	43*	48*	48*	62+	66	96	100

Snn, UM, Fnn are speech rate levels: S: slowed down; UM: unmodified; F: speeded up; 'nn' : compression/expansion rate in percentage.  
'\*' means scores are within the 'comfortable range' (scores between 40 and 60) and '+', nearly in the comfortable range (scores between 35 and 65).



## 6.2.2. Experiment 1-B

### 6.2.2.1. Method

#### *1. Design*

A factorial 2 x 5 design was used with two speakers (KS and TK) and five different speech rates (change rate of segmental durations of the source utterances: 0.8, 0.9, 1.0, 1.1 and 1.2).

#### *2. Speech materials and stimulus preparation*

The utterances were politely spoken by a slower speaker KS (speech rate: 10.8 mora per sec) and by a faster speaker TK (speech rate: 11.8 mora per sec). Rate was realised by linearly compressing or expanding each segmental duration by means of computer resynthesis based on the TD-PSOLA technique. The change rates in segmental duration were 0.8, 0.9, 1.0, 1.1 and 1.2. As was discussed in Section 3.5.1.3, this linear change has been a concern because of potential artefacts introduced by this manipulation (e.g., Apple *et al.*, 1979; Bruce Brown, 1980). Unfortunately no satisfactory rules for predicting segmental durational changes in function of speech rate in Japanese exist at the present time. Therefore this linear algorithm was used to implement durational changes.

Some degradation of speech may be caused by the computer manipulation. Apple *et al.* (1979) mentioned that some quality variation was apparent across different speakers in LP resynthesised speech, in terms of nasality for example. This degradation could cause some undesirable artefact in politeness ratings because of the close link between politeness and naturalness perception. This question is addressed in Experiment 3 (Section 6.3). In order to preserve the factor of



degradation caused by the PSOLA manipulation at the same level as that for the other rate versions, all the unmodified versions were resynthesised with a rate change factor of 1.0 .

### *3. Rating sessions*

The politeness ratings were those already reported in Section 6.1.1, that is with 164 stimuli (6 occurrences each of the 26 conditions plus 8 dummies). As listener variables were considered to be important, speech rate of the subjects was assessed at the end of the session; subjects were presented written text in Japanese meaning "Hello, how do you do. I'm afraid as I haven't brought anything to write with, could I borrow the ball-point pen over there?", and asked to speak them to the experimenter twice as naturally as possible, in front of a small microphone on the desk. The written text given to the subjects is included in Appendix E.

#### **6.2.2.2. Results and discussion**

The politeness scores were obtained by measuring the distance between subjects' markings from a mid point on the bipolar scale. Scores could range between -4 (very impolite) and +4 (very polite). Kendall's coefficient of concordance (W) was calculated to assess the general agreement among 20 listener-judges' ratings of the five different rate versions. The mean reliability assessed by the Spearman rank-order correlation coefficient (Mean  $r_s$ ) is also reported with W.  $W = 0.47$  (Mean  $r_s = 0.44$ ,  $N = 20$ ) for KS's utterances and  $0.40$  (Mean  $r_s = 0.37$ ,  $N = 20$ ) for TK's utterances ( $p_s < 0.0001$ ). The effective reliability was very high ( $R > 0.9$ ). The intra-judge agreement was assessed by calculating the significance of the factor of each subject's six repetition scores for each condition. Two ANOVA tests were carried out with factors of rate (five rates) and subjects' repetition factor (six repetitions) for utterances of both speakers separately. The repetition factor was

non-significant for KS's utterances ( $p = 0.128$ ), but came closely to the significance level of 0.05 for TK's utterances ( $p = 0.052$ ). This result shows that the subjects were consistent in rating for KS's versions while they had some difficulty in judging TK's different rate versions.

An ANOVA test was performed with factors of speakers (two speakers) and speech rates (five rates). The results of this ANOVA test are attached in Appendix 2 (1-B-1), and the significant effects at the level of 0.05 or better are summarised in Table 6.7.

TABLE 6.7. ANOVA results of Experiment 1-B: significant effects at the level of 0.05 or better.

	<i>F</i>	<i>p: significance of F</i>	<i>eta-squared (%)</i>
<u>MAIN EFFECTS</u>			
Speaker*1	23.73	**	28.5
Rate*2	11.99	**	15.7
<u>INTERACTIONS</u>			
Speaker and Rate*2	7.88	**	2.4

\*1  $df_{effect} = 1$  and  $df_{error} = 19$

\*2  $df_{effect} = 4$  and  $df_{error} = 76$

\*\* :  $p < 0.001$

The ANOVA test showed significant main effects of speaker and rate, and a significant interaction between these two main factors ( $ps < 0.05$ ). Eta-squared, as an indicator of the weight of contribution of the factors, showed that the main effects were stronger than the interaction, while the speaker factor was more salient than the rate factor. The mean values of politeness scores across 20 subjects for the



five different versions of the sentence are shown in Table 6.8. The politeness scores for KS and TK are shown separately in Fig. 6.13. Both curves show an inverted-U shape as a function of rate. However, it is noticed that the curve for TK is slightly shifted to the faster rates compared with the KS curve. This difference in speech rate for the maximum politeness scores agrees with the difference in the comfortable speech rates for both speakers: the comfortable range in the listening test in the previous section was 11 ~ 12 mora per sec for KS and 12 ~ 13 mora per sec for TK (see Table 6.6).

TABLE 6.8. Mean politeness ratings with standard deviations (SD) in Experiment 1-B.

<i>Speaker</i>	<i>Speech rate level</i>	<i>Speech rate (mora/sec)</i>	<i>Mean</i>	<i>SD</i>
KS	S20	8.9	-1.85	1.279
	S10	9.8	-1.21	1.431
	UM	10.8	-0.31	1.260
	F10	11.9	0.03	1.298
	F20	13.5	-0.52	1.249
TK	S20	9.8	0.13	1.238
	S10	10.6	0.64	1.046
	UM	11.8	1.39	1.019
	F10	13.0	1.33	0.934
	F20	14.7	0.26	1.156

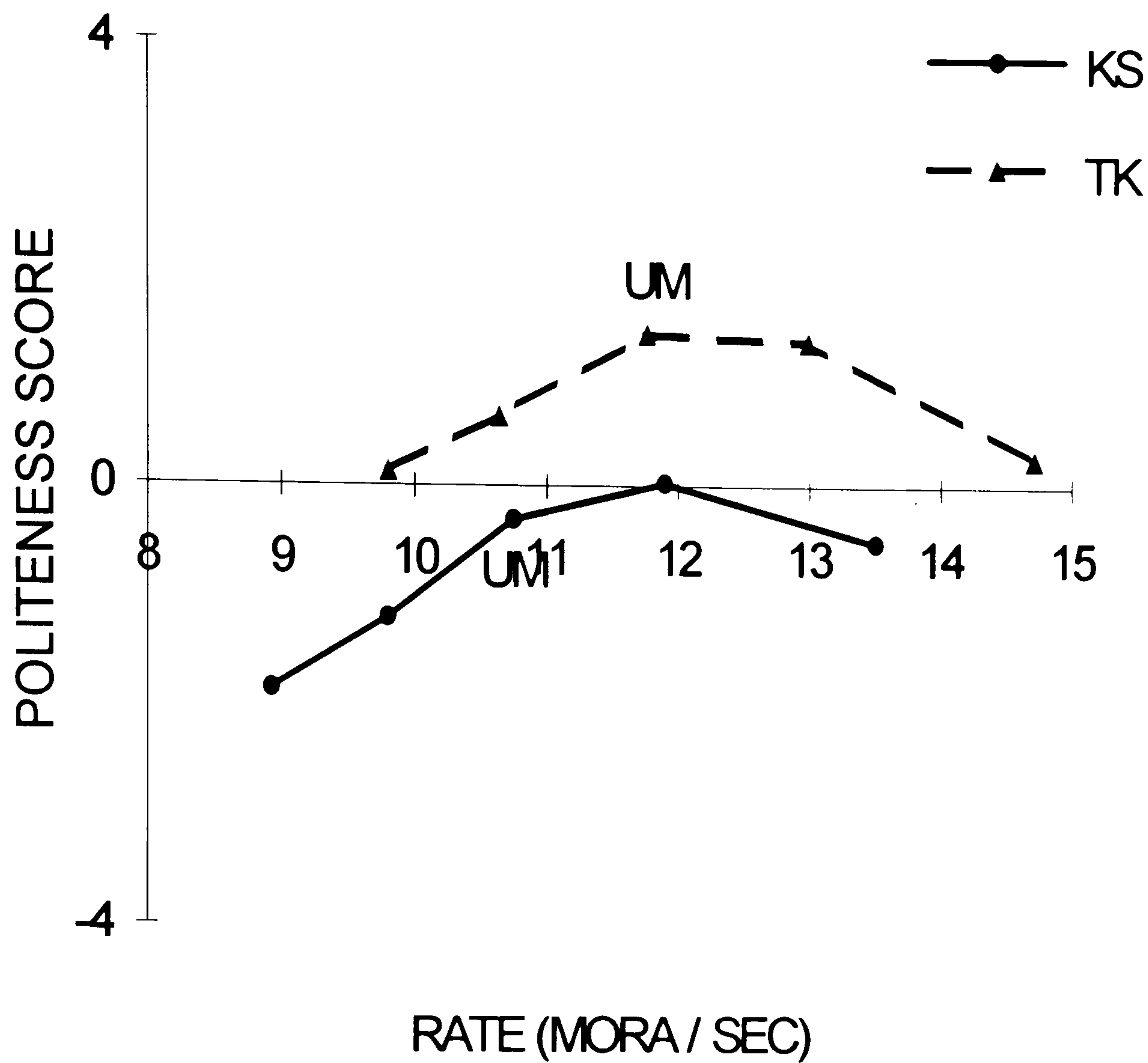


FIG. 6.13. Mean values of politeness scores across 20 subjects for five different rate versions of the sentence originally spoken by KS and TK in Experiment 1-B. The unmodified speech rate (UM) is marked on each function.

Although a significant main effect of rate was obtained, the consistency among subjects' judgements was not very high. So the factors of the speech rate of the subjects and the accent of the subjects were examined in relation to individual rate preference. No significant accent effect was found while an interesting relationship between the speech rate of the subjects and their rate preference emerged, as detailed below.

The speech rate of the subjects was assessed by the measurement of the speaking time of the utterance (exclusive of pauses) recorded at the end of the rating sessions. The speech rates of each subject are shown in Table 6.9.



TABLE 6.9. Speech rates of subjects.

<i>Subject</i>		<i>Speech rate</i> <i>(mora/sec)</i>		<i>Rate Category</i>
<i>Sex</i>	<i>ID</i>	<i>Trial 1</i>	<i>Trial 2</i>	
Male	1	error	10.6	FAST
	2	9.9	10.1	FAST
	3	error	10.6	FAST
	4	9.9	8.6	SLOW
	5	8.6	9.3	MIDDLE
	6	9.3	9.4	MIDDLE
	7	8.4	9.3	MIDDLE
	8	10.5	10.1	FAST
	9	8.6	8.6	SLOW
	10	9.8	9.7	MIDDLE
	11	10.8	error	FAST
	12	error	11.0	FAST
Female	1	error	8.2	SLOW
	2	8.9	8.9	MIDDLE
	3	10.0	9.9	FAST
	4	9.0	9.0	MIDDLE
	5	8.1	error	SLOW
	6	8.3	8.3	SLOW
	7	9.0	8.9	MIDDLE
	8	8.2	8.2	SLOW

Each subject spoke the utterances twice. Since the first trial was not very smoothly spoken in most cases, utterances in the second trial were used. When the second trial was not successful (e.g., tongue twist, wrong sentence), the first utterance was used. The speech rates of the 20 subjects showed a significant sex difference: the female subjects ( $N = 8$ ) spoke significantly more slowly than the male subjects (2-tailed t-test,  $p < 0.01$ ). It is interesting to speculate that this sex difference in rate might reflect a social expectation that women should be more polite than men, as well as a social expectation that slower speech is associated with politeness as Ogino and Hong's (1992) survey showed. The speech rate of the subjects were categorised into three groups, for male and female subjects separately, by using a cluster analysis: slow-speaker (2 male and 4 female), middle-speaker (4 male and 3 female) and fast-speaker (6 male and 1 female) (see 'Rate Category' in Table 6.9). An ANOVA with factors of speech rate of listener and sex of listener as the between-subject factors, and rate of utterance and speaker of utterance as the within-subject factors, showed significant main effects of speaker and rate, and significant interactions between speech rate of listener and rate of utterance and a significant three-way interaction between speech rate of listener, sex of listener, and rate of utterance ( $ps < 0.001$ ). The results of this ANOVA test is attached in Appendix 2 (1-B-2). The important effects are summarised in Fig. 6.14, which illustrates an interesting relationship about the interaction between the speech rate of the 12 male subjects and their rate preference. Slow speakers rated slower utterances as more polite than middle/fast speakers, while the middle/fast speakers preferred faster versions. The female subjects' data, however, showed no clear difference between the slower-speaker group and the faster-speaker group (Fig. 6.15).



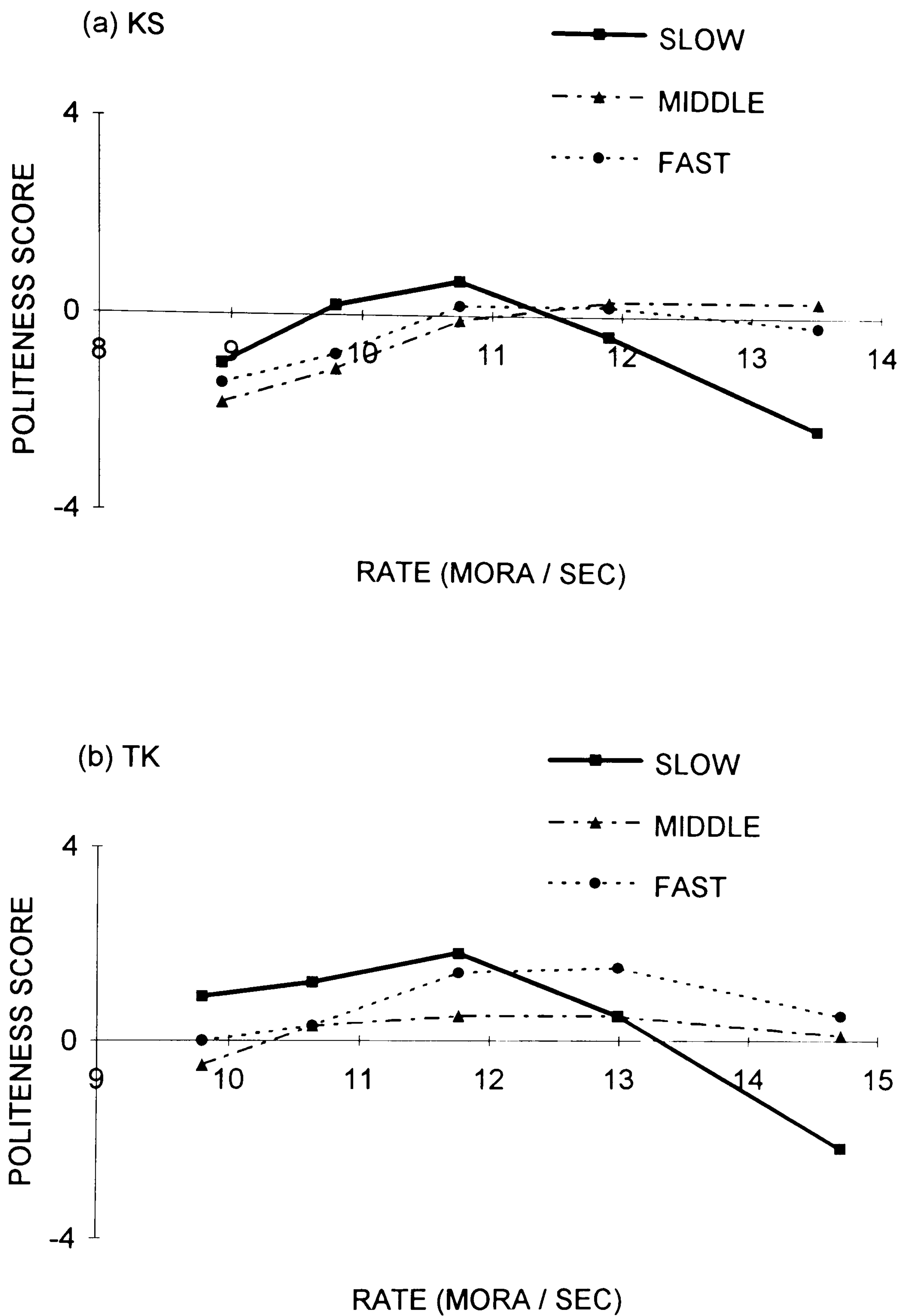


FIG. 6.14. Rate preferences of 12 male subjects in Experiment 1-B as a function of the rate of the utterances. SLOW, MIDDLE and FAST are the category of subjects in terms of their speech rate. The politeness scores for the five different rate versions of the sentence originally spoken by KS (a) and those for the utterances by TK (b) are shown separately.

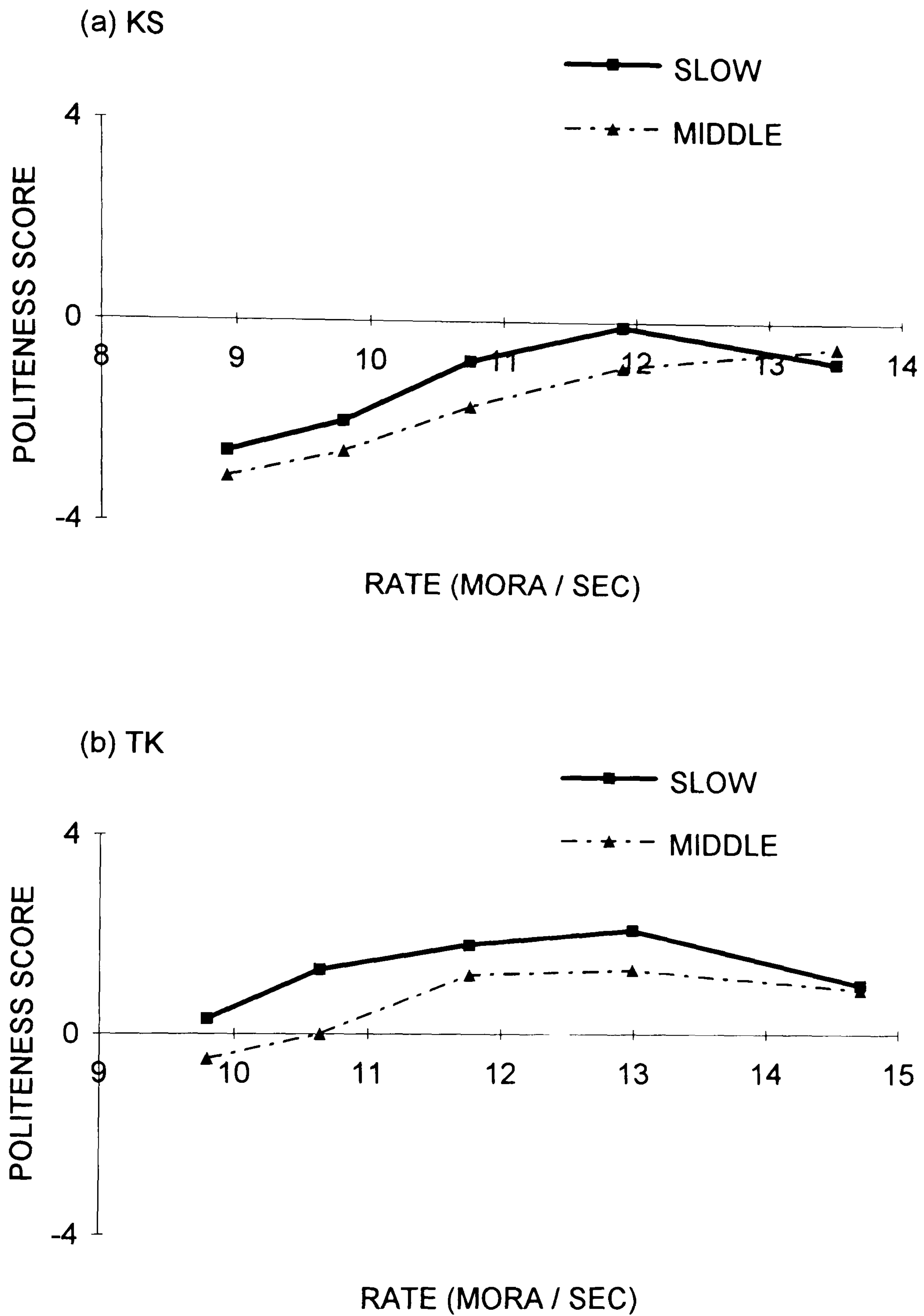


FIG. 6.15. Rate preferences of 7 female subjects in Experiment 1-B as a function of the rate of the utterances. SLOW and MIDDLE are the category of subjects in terms of their speech rate. Since only one female subject was categorised as FAST, the data was excluded from this figure. The politeness scores for the five different rate versions of the sentence originally spoken by KS (a) and those for the utterances by TK (b) are shown separately.



### 6.3. Experiment 3: The role of perceived naturalness

In the previous experiment it was found that listener characteristics influenced their judgements on politeness. For example, subjects who were slower speakers themselves preferred slower utterances whereas faster speakers preferred faster utterances. People appear to judge utterances which sound more 'natural' to them as more polite unless the utterances are originally distinctively impolite. (The word 'natural' is used broadly here, to include features such as perceived appropriateness or comfortableness.) This tendency suggests the importance of perceived naturalness in politeness judgements, therefore, the relationship between politeness and naturalness is the focus of this experiment. However, since the term 'naturalness' is again a broad concept like 'politeness', it includes several different aspects (e.g., abnormality in terms of articulation, prosody and voice quality, and appropriateness in specific situations, etc.). Since all the stimulus utterances used in Experiment 3 were created using human utterances, and did not sound 'abnormal' in terms of speech quality, the focus here was, therefore, on the aspect of appropriateness.

#### 6.3.1. Method

##### *1. Speech material*

The same material as that for the experiment on speech rate (Experiment 1-B) was used: the 'luggage' sentence spoken by the two speakers, KS and TK, with five different rates (change rate of segmental duration of the source utterance: 0.8, 0.9, 1.0, 1.1 and 1.2) manipulated by means of the TD-PSOLA technique.

##### *2. Stimulus presentation and two alternative forced-choice procedure*

A two alternative forced-choice (or paired comparison) method for the 10 conditions (i.e., 2 speakers x 5 rates) was used to measure perceived politeness and perceived naturalness (Watkins and Makin, 1994). This paired comparison method has both advantages and disadvantages. One of the disadvantages is that this type of judgement is very artificial in the sense that people do not usually assess politeness in this way. However, the great advantage is that it is easier to make judgements and the scores can be very reliable compared with a rating scale method. This is mainly because subjects have a reference in each judgement. This method was adopted for the present experiment, instead of a rating scale method which was used in the previous experiments, because naturalness rating on a scale was found to be very difficult for subjects in the former studies. This is especially so when all the stimuli are not very different from each other in terms of naturalness. In the rating scale method subjects have to map each utterance on a linear scale without reference utterances for 'very unnatural' and 'very natural'. However, subjects tend to lose their sensitivity to naturalness quickly when they hear a number of similar-sounding stimuli many times, and therefore the scores tend to fluctuate. This fluctuation could be greater than the differences between stimulus conditions. In fact, there was no significant difference in naturalness ratings for four conditions based on the 'angry' speaking style of the first part of utterances in Experiment 2 (Section 6.1.2).

First, politeness scores were collected. On each trial, subjects heard two utterances successively, and selected which utterance sounded more polite to them. Each utterance was compared with every other utterance, using both orders of presentation. Subjects were presented with a total of 458 trials, consisting of 5 repetitions of 90 trial types and 8 dummy trials, randomised differently for each subject, over two days. Following a practice session, one session consisting of 38 trials and three sessions of 60 trials were run. The first trial of a session was a dummy and there was a short break between sessions. Before the first session



started, subjects were given short written instructions (Appendix H). Subjects were asked to judge politeness in a specific situation in which a young customs officer was speaking to a respectable gentleman. They were not told that they were going to judge naturalness later when they started sessions for politeness scores. The two sessions each day lasted about 30 minutes. Two daily sessions for naturalness were next run, again with a total of 458 trials, following exactly the same procedure as that for the politeness sessions. Before the naturalness session started, subjects were given written instructions, which is also attached in Appendix H, and asked to judge naturalness as a global judgement, giving consideration to such various factors as speech quality, tempo, the way of speaking and the appropriateness in a given situation

### *3. Subjects*

There were four paid subjects (2 male and 2 female). All subjects were native speakers of Japanese, ranging in age between 27 and 31 and were postgraduates at British universities.

### **6.3.2. Results and discussion**

The politeness/naturalness scores were calculated as the number of times an utterance was judged more polite/natural in a comparison, divided by the total number of occurrences of each utterance. The scores could range from 0 (most impolite/unnatural) to 1 (most polite/natural). All the scores for utterances by both speakers are shown in Table I.1 in Appendix I. The scores for TK's utterances by the four subjects are shown individually in Figs. 6.16 to 6.19.

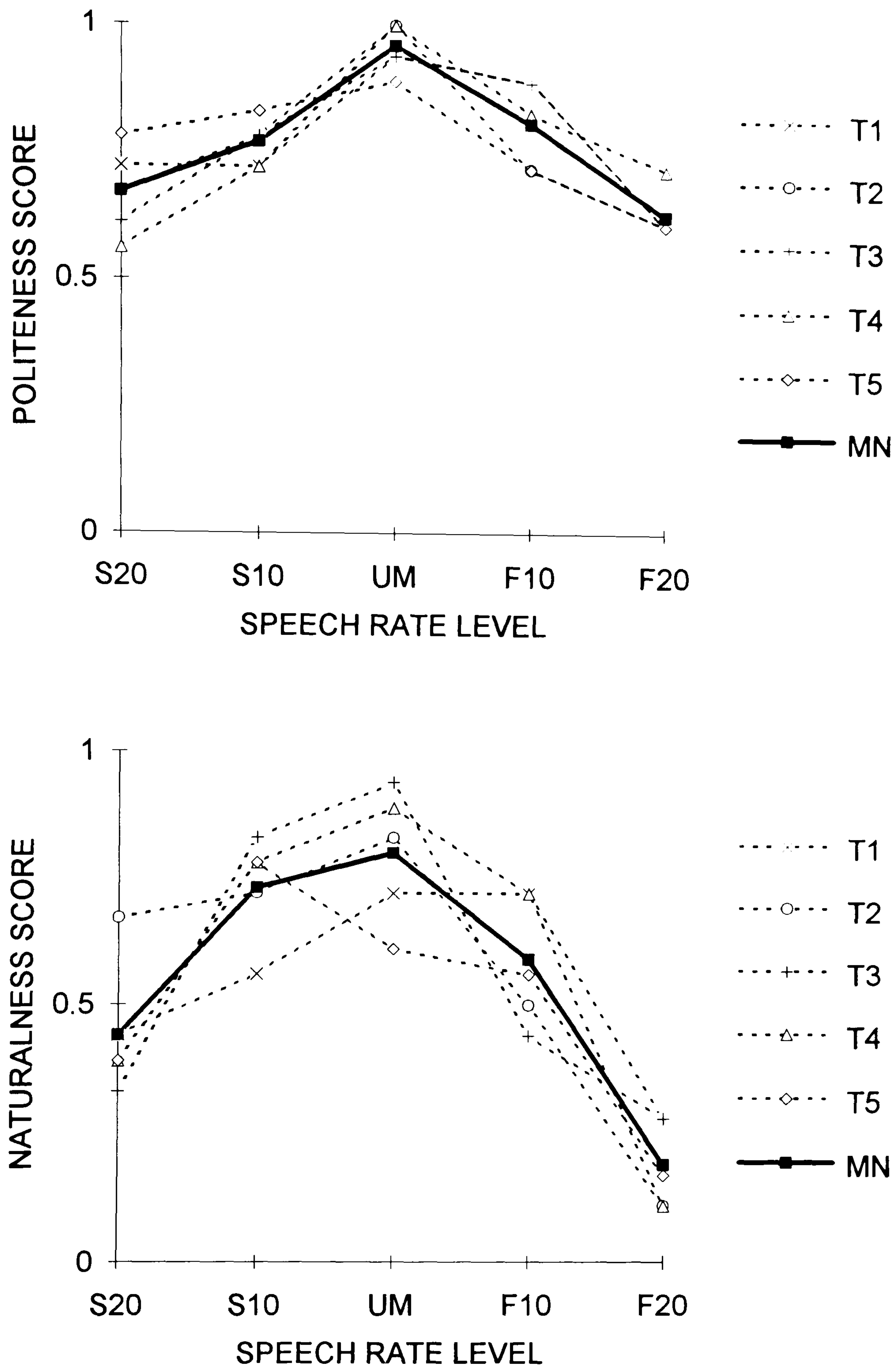


FIG. 6.16. A male subject (M1)'s politeness scores and naturalness scores for the utterances by TK in Experiment 3; scores for each trial block (T1 ~ T5) and the mean values (MN) across 5 trial blocks are shown. 'Speech rate level' is a level of compression/expansion rate in segmental duration of the source utterance: S20 is 20% expansion, S10, 10% expansion, UM, unmodified duration, F10, 10% compression and F20, 20% compression.



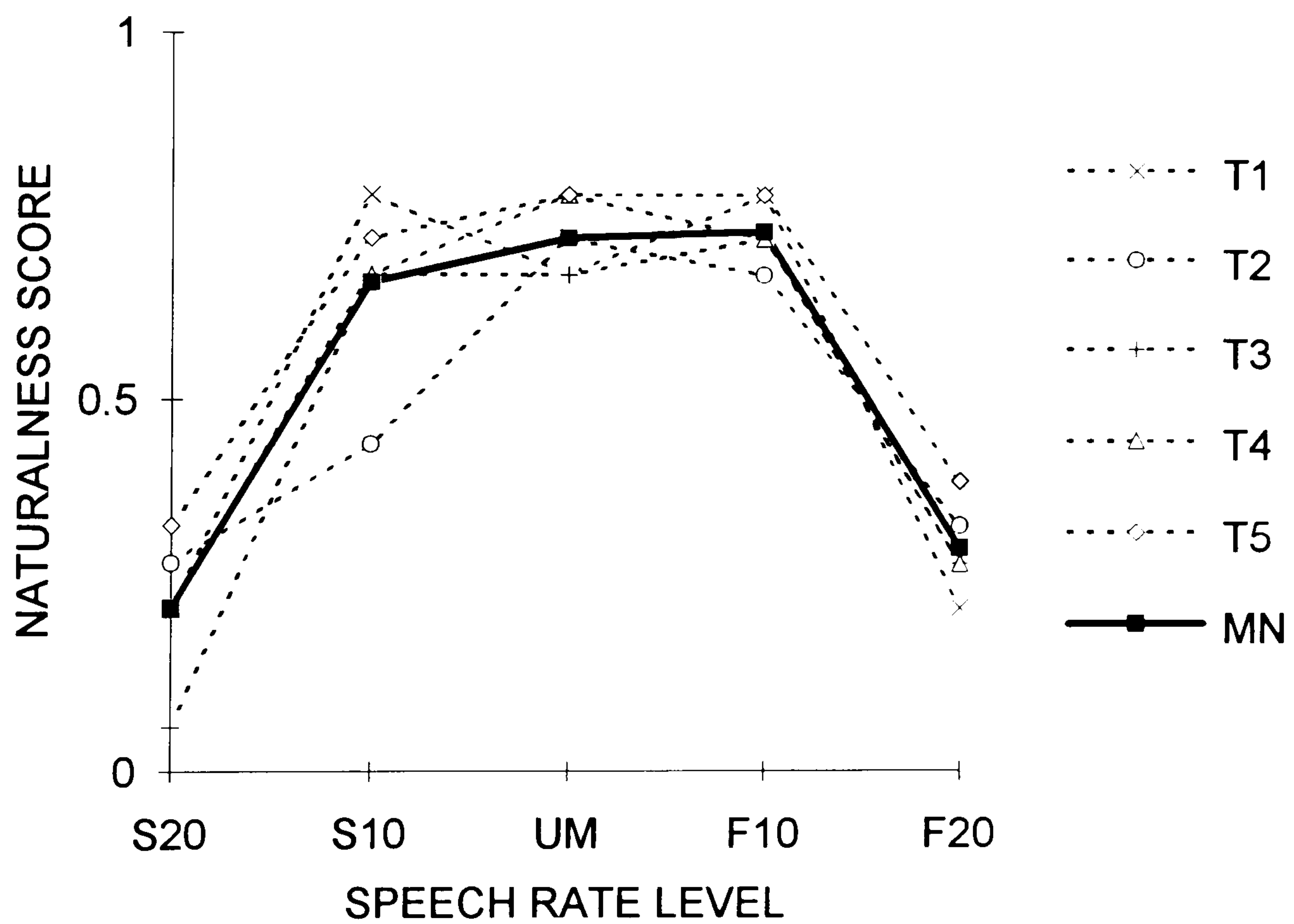
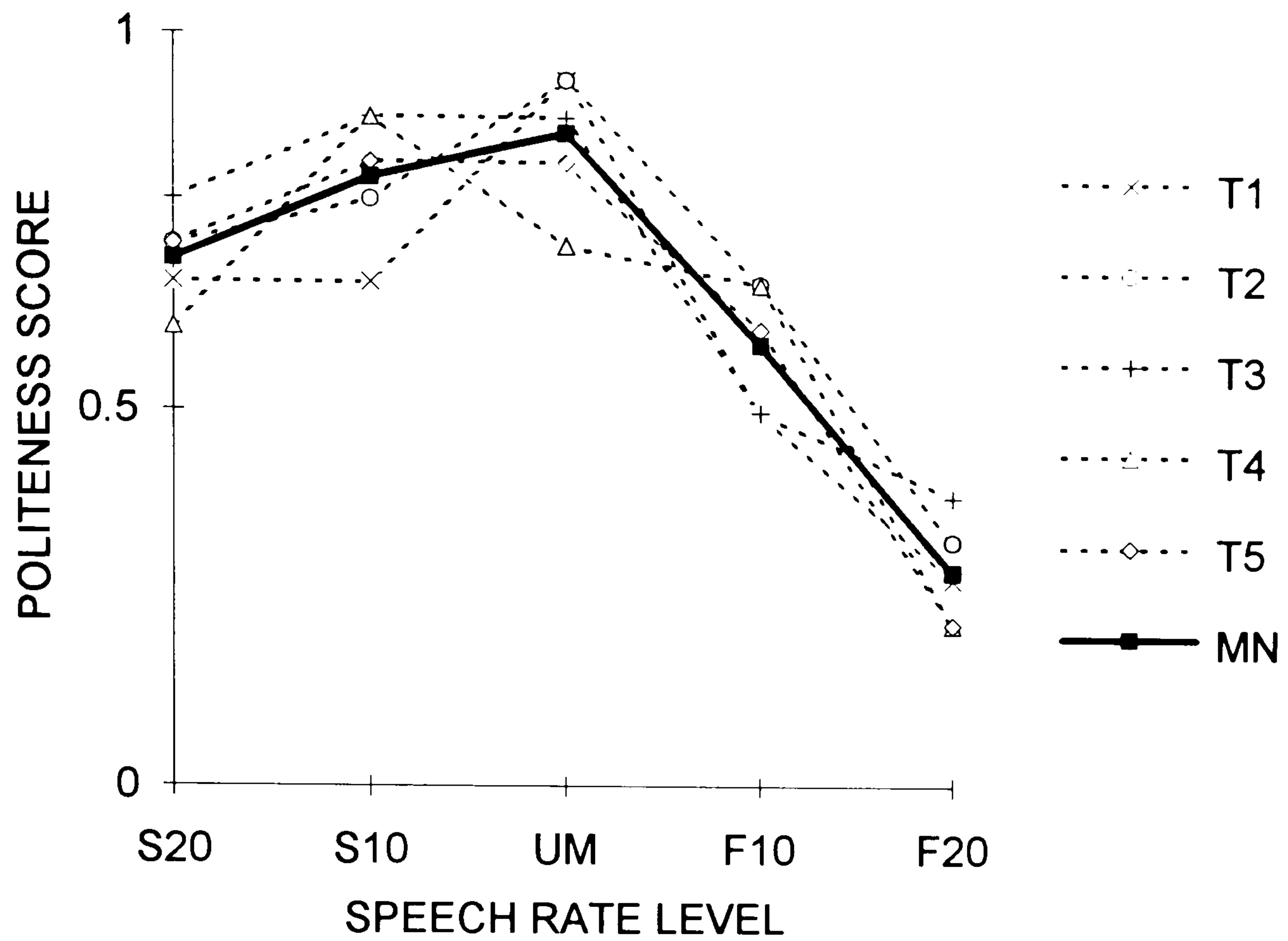


FIG. 6.17. A male subject (M2)'s politeness scores and naturalness scores for the utterances by TK in Experiment 3; scores for each trial block (T1 ~ T5) and the mean values (MN) across 5 trial blocks are shown. 'Speech rate level' is a level of compression/expansion rate in segmental duration of the source utterance: S20 is 20% expansion, S10, 10% expansion, UM, unmodified duration, F10, 10% compression and F20, 20% compression.

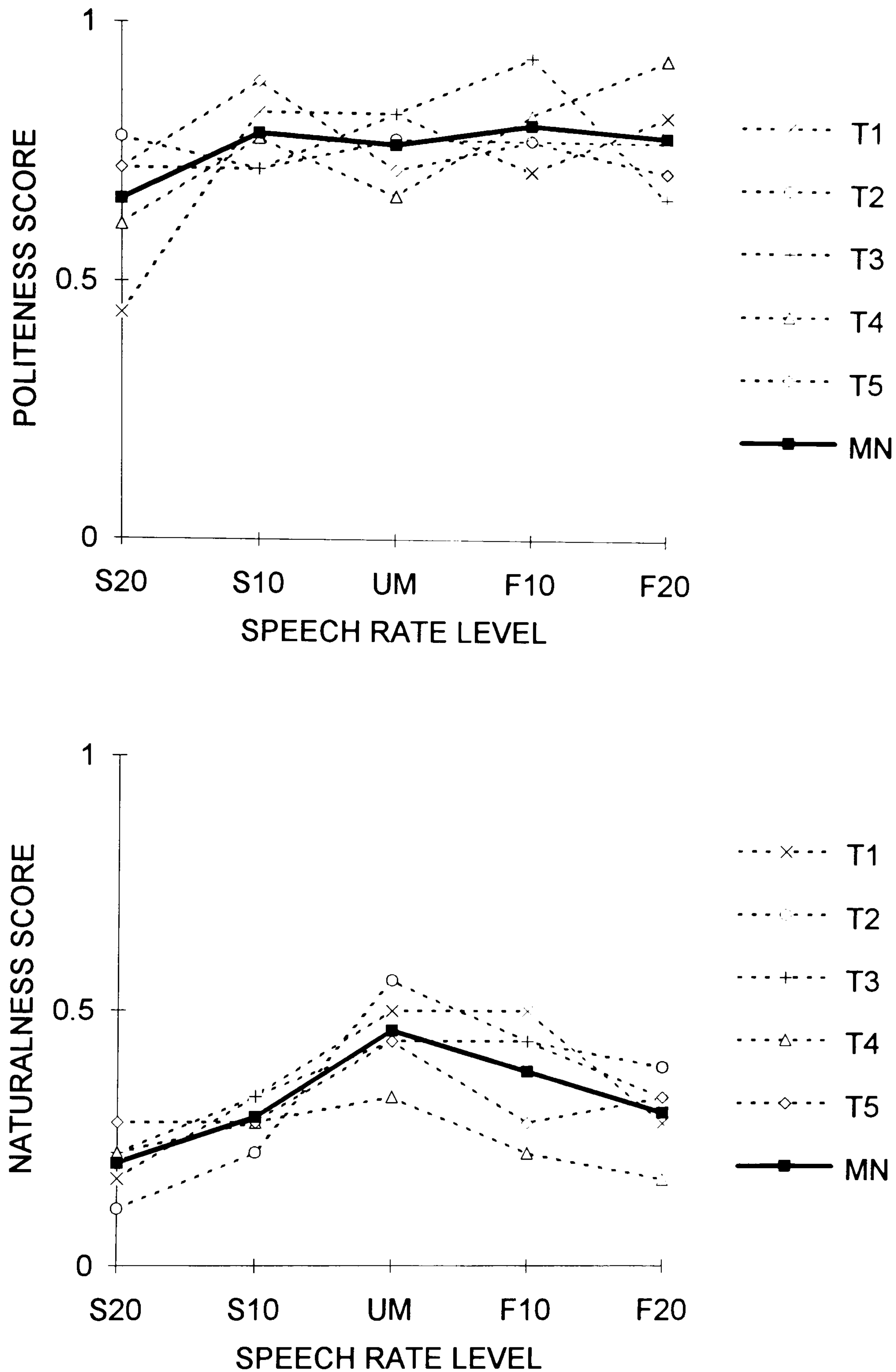


FIG. 6.18. A female subject (F1)'s politeness scores and naturalness scores for the utterances by TK in Experiment 3; scores for each trial block (T1 ~ T5) and the mean values (MN) across 5 trial blocks are shown. 'Speech rate level' is a level of compression/expansion rate in segmental duration of the source utterance: S20 is 20% expansion, S10, 10% expansion, UM, unmodified duration, F10, 10% compression and F20, 20% compression.



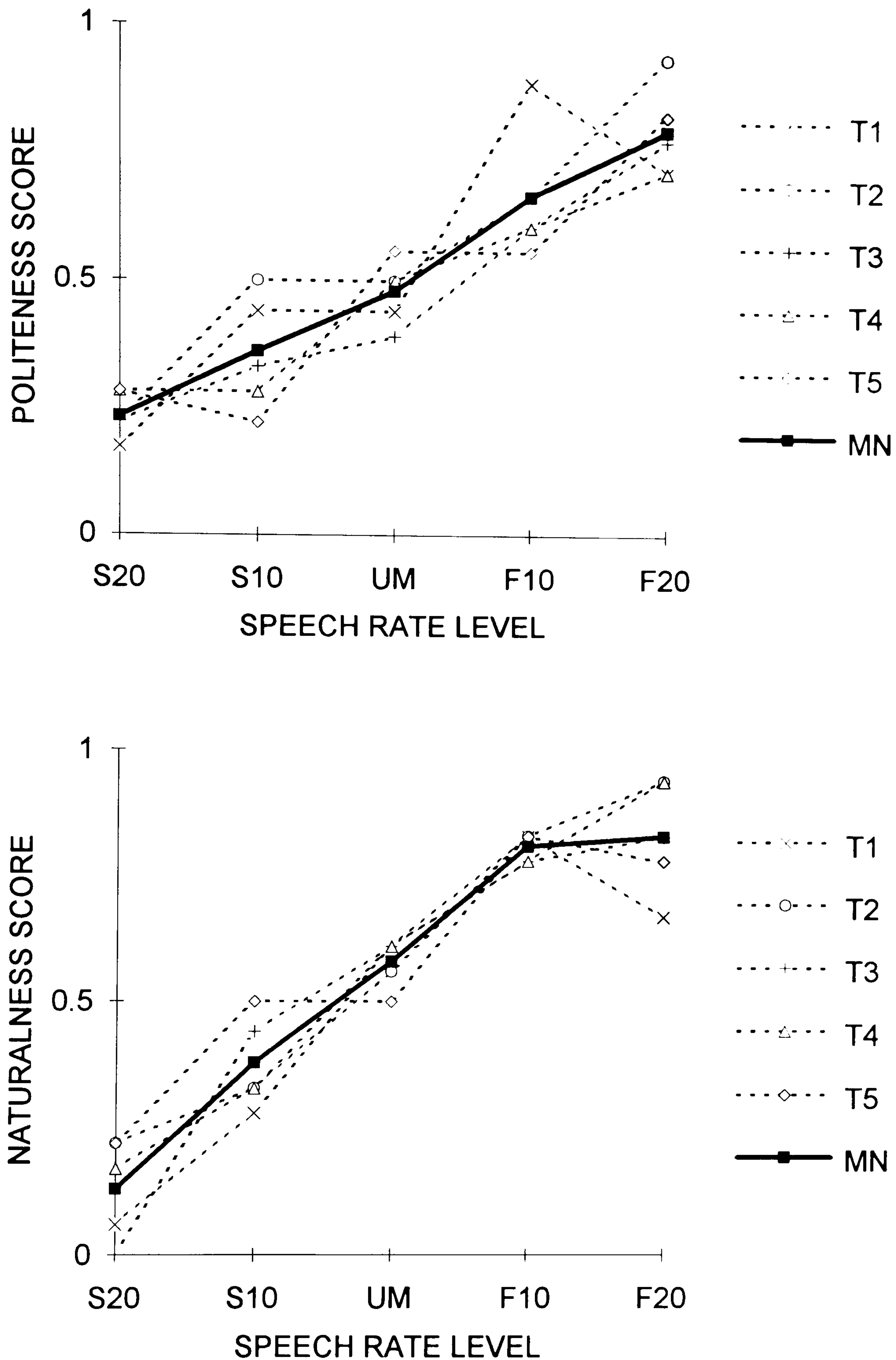


FIG. 6.19. A female subject (F2)'s politeness scores and naturalness scores for the utterances by TK in Experiment 3; scores for each trial block (T1 ~ T5) and the mean values (MN) across 5 trial blocks are shown. 'Speech rate level' is a level of compression/expansion rate in segmental duration of the source utterance: S20 is 20% expansion, S10, 10% expansion, UM, unmodified duration, F10, 10% compression and F20, 20% compression.

Kendall's coefficient of concordance ( $W$ ) was calculated to assess inter-set agreement among each subject's five sets of judgements for the five different rate versions for politeness and naturalness (Table 6.10). This, together with Figs. 6.16 and 6.19, shows that there is a very high level of consistency between the 5-set trials of judgements by all the subjects, except Subject F1's judgements of politeness for TK's utterances (Fig. 6.18). Both politeness scores and naturalness scores are found to have a curvilinear relationship with speech rate of utterances (Fig 6.20), and the relationship between the politeness scores and naturalness scores seem to be linear when scores for KS's utterances and those of TK were examined separately (Fig. 6.21). Therefore, the Pearson product-moment correlation coefficient ( $r$ ) was calculated to assess the correlation between them. The correlation coefficients between the mean values for politeness scores of four subjects and those for naturalness scores, and the correlation coefficients at the individual level are shown in Table 6.11, all of which show a fairly high to very high level of positive correlation.

TABLE 6.10. Inter-trial agreement among each subject's five trial blocks for the five different rate versions of the sentence originally spoken by the two speakers (KS and TK) for politeness and naturalness in Experiment 3, using Kendall's coefficient of concordance ( $W$ ) with a level of significance better than 0.01.

<i>Speaker of utterance</i>	<i>Subject</i>	<i>Kendall's <math>W</math> for politeness scores</i>	<i>Kendall's <math>W</math> for naturalness scores</i>
KS	M1	0.77	0.97
	M2	0.83	0.84
	F1	0.74	0.86
	F2	0.97	0.95
TK	M1	0.80	0.86
	M2	0.92	0.87
	F1	0.19 (*1)	0.66
	F2	0.89	0.94

\*1: Not significant ( $p = 0.44$ )



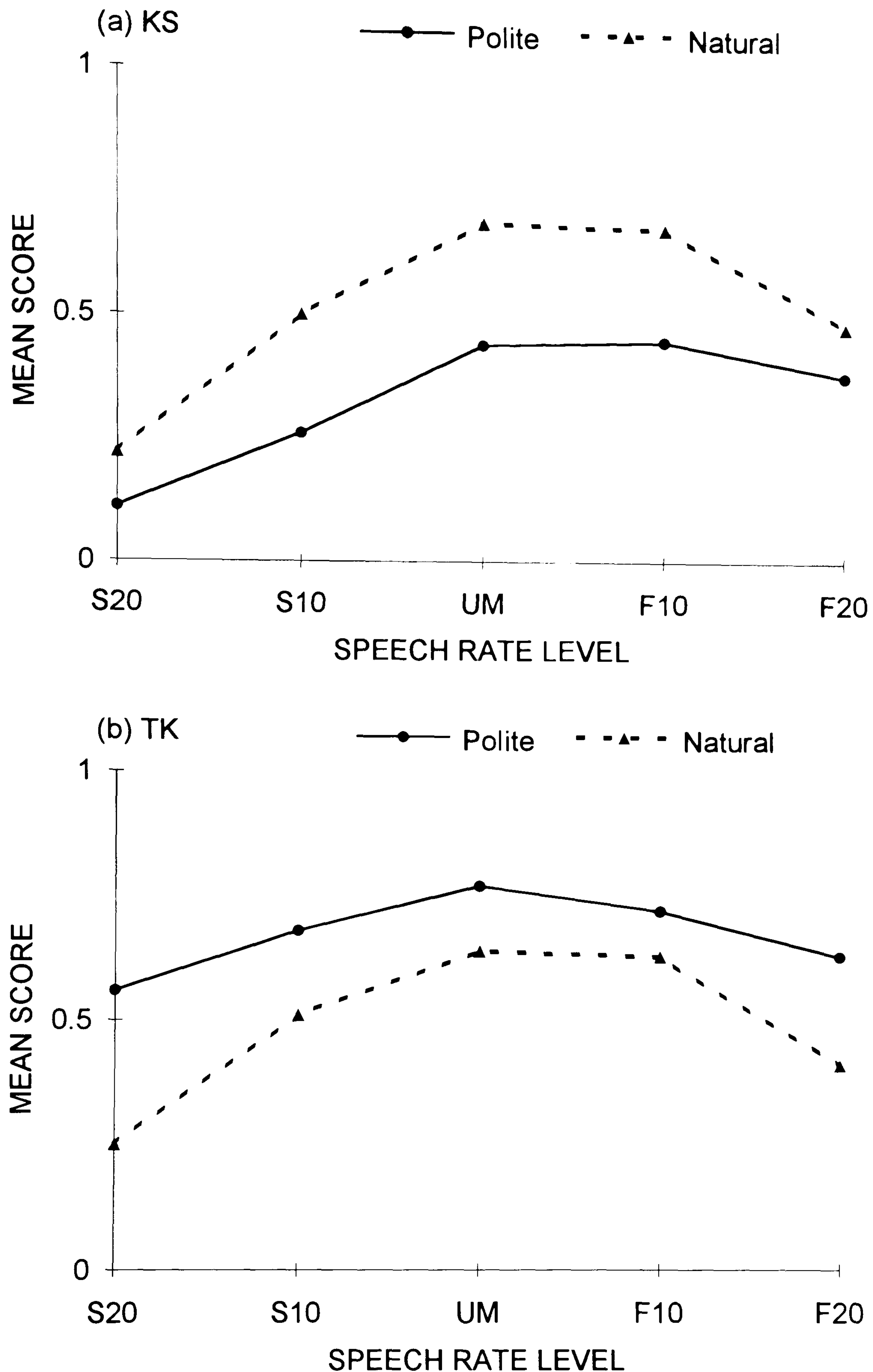


FIG. 6.20. Mean politeness and naturalness scores for the four subjects for the five different rate versions in Experiment 3; scores for the utterances by KS (a) and those for the utterances by TK (b) are shown separately. 'Speech rate level' is a level of compression/expansion rate in segmental duration of the source utterance: S20 is 20% expansion, S10, 10% expansion, UM, unmodified duration, F10, 10% compression and F20, 20% compression.

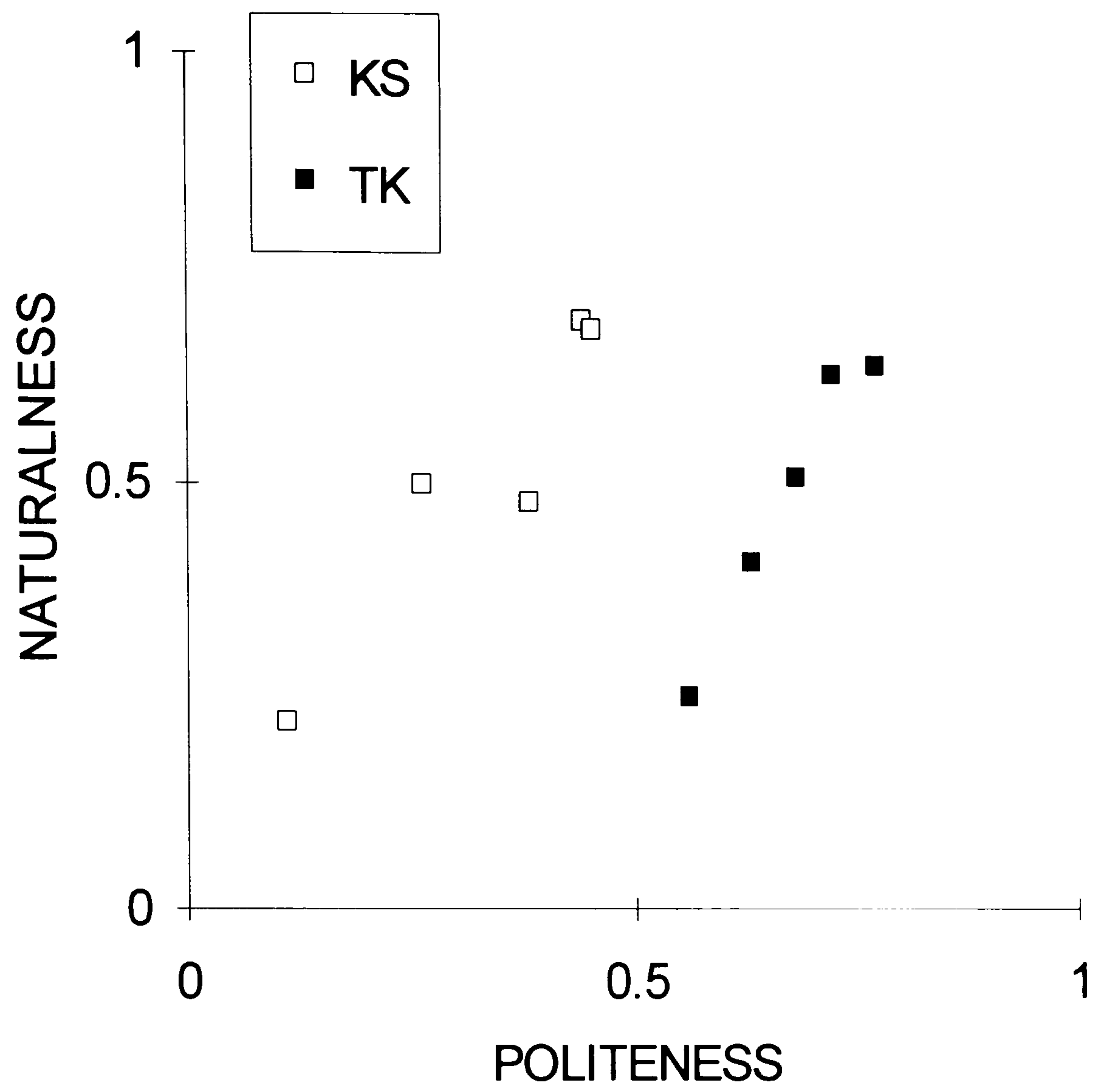


FIG. 6.21. Mean politeness and naturalness scores for the four subjects in Experiment 3; the five different rate versions of the sentence originally spoken by KS and TK are shown separately.



TABLE 6.11. Correlations between politeness scores and naturalness scores of the five different rate versions of the sentence originally spoken by the two speakers (KS and TK) by the Pearson product-moment correlation coefficient ( $r$ ) in Experiment 3.

<i>Speaker of utterance</i>	<i>Subject</i>	<i>Pearson's <math>r</math></i>
KS	M1	0.798
	M2	0.607
	F1	0.626
	F2	0.991
	Grand Mean	0.939
TK	M1	0.861
	M2	0.521
	F1	0.623
	F2	0.970
	Grand Mean	0.981

Although politeness and naturalness are found to be correlated among different rate versions of the same source utterance, they are clearly not the same thing. For example, Subject M2 perceived slower versions less natural, but more polite than the faster versions (Fig. 6.17). An important point is that all the utterances used as stimuli in this experiment at least did not sound impolite: according to the politeness ratings of these source utterances in Experiment 1-B, KS's actual utterance had a politeness score of -0.31 out of -4 (very impolite), and TK's actual utterance had a score of +1.39 out of +4 (very polite). So the politeness levels of the present stimuli fall within a limited degree of politeness, that is between fairly impolite and quite polite. It is natural to suppose that both perceived naturalness and perceived politeness each moderate the other: over polite utterances may sound rather

unnatural, and natural impolite utterances may sound more impolite than unnatural impolite utterances because the former are more realistic.

The positive correlation between speech rate of listeners and their preferred rate of utterances, which was examined in Experiment 1-B, was informally supported by Subject F2's politeness scores (Fig. 6.19), which showed strong preference for faster utterances with F2 herself being a very fast speaker.

#### **6.4. Summary**

Acoustic analysis, which is described in Chapter 5, showed that global rate of articulation and f0 movement of the final vowel of the sentences were adopted consistently by all the speakers to differentiate polite and non-polite utterances. On the other hand, f0 related variables such as f0 level, f0 range and f0 rate of change were not used consistently across these six speakers. The use of these f0 variables was different even across utterances of two sentences spoken by the same speaker in some cases. In order to observe the effects of these variables which were used differently according to the speaking styles, final f0 movement (Experiments 1-A and 2) and speech rate of utterance (Experiment 1-B) were conducted by using digital resynthesis.

Experiments 1-A and 2, which examined the effects of final f0 movement, showed that the prosody of the final vowel of the sentence had a great impact on politeness judgements: prosody information through the last 100 ms or so changed the overall impression of the utterance. Final durations which were not too long nor too short, and final rises were rated more positively than very long or short final durations and final falls. The preference of a rising tone may be related to the fact that the default tone of the sentence is a rising one. Experiment 2, whose aim was to examine the effects of speaking style, confirmed that the speaking style of the final



mora was also very influential in politeness judgements. It was also found that there was an interesting relationship between accent of listener and style preference (i.e., polite and casual): subjects from the western part of Japan rated the speakers' 'casual' style more polite than the 'polite' style, while subjects from the eastern part rated them as the speakers intended. There was no significant difference in their preference for final prosody.

Experiment 1-B, which focused on the effects of speech rate of utterance, showed that the main factors of speaker and speech rate, and the interaction between them were significant and the function relating politeness and speech rate was of an inverted-U shape. However, the rate factor was not powerful enough to override the speaker/style difference in terms of politeness. This may be due to the fact that the rate change used in this experiment was achieved by linear compression or expansion of each segmental duration, which does not usually take place in tempo altering by human speakers. Another interesting finding was the positive correlation between speech rate of listener and their preferred rate of utterance, which confirmed the importance of the listener factor in perceptual tasks.

In Experiment 3 politeness scores for different rate variations of the polite source utterances were examined in relation to perceived naturalness by using a paired comparison method. The results showed that there was a very high positive correlation between relative politeness and relative naturalness among the five different rate versions of the same utterance. As far as changes in rate are concerned, naturalness appears to be an important factor. Naturalness influences politeness judgements, and is probably implicitly influenced by politeness too.

# CHAPTER 7

## GENERAL CONCLUSIONS

### 7.1. Summary of the findings

The present study has investigated prosodic features for signalling politeness in Japanese. Samples of polite and non-polite utterances of two routine question sentences were collected from simulations by six male native speakers with a role play method. The recordings were later digitised and acoustically analysed with the specific focus on  $f_0$  and temporal features. Acoustic analysis showed that global rate of articulation and  $f_0$  movement of the final vowel of the sentences were used differently by all the speakers for polite and non-polite situations, while  $f_0$  level,  $f_0$  range and  $f_0$  rate of change were not used consistently across these six speakers. In order to confirm the effects of these variables which were used differently according to the speaking styles, final  $f_0$  movement (Experiments 1-A and 2) and speech rate of utterance (Experiment 1-B) were systematically manipulated by using digital resynthesis.

In Experiments 1-A and 2 the importance of the final part of the utterance was confirmed. The impact of the speaking style and prosody (i.e., duration and  $f_0$  direction) was so great that the changes of the last 100 ms or so of the utterance changed the overall impression of the utterances. Final durations which were not too long or too short, and final rises were rated more positively. The final  $f_0$  direction, however, appears to be closely related to 'unmarkedness' (i.e., the default tones of the expression).

In Experiment 1-B the role of speech rate, variations of which were realised by linear changes in segmental durations, was investigated. The rate factor was found



to be significant, and the function relating politeness and speech rate was of an inverted-U shape (i.e., the normal rate being rated as the most polite). However, this rate factor was not strong enough to override the speaker/style differences. There was a positive correlation between speech rate of listeners and their preferred rate of utterances.

Experiment 3 was carried out to examine the effects of perceived naturalness on politeness judgements with different speech rate variations of a polite utterance by using a paired comparison method. There was a very high positive correlation between relative politeness and relative naturalness among the five different rate versions of the same utterance. As far as changes in rate are concerned, naturalness appears to influence politeness judgements, and is probably implicitly influenced by politeness too.

In summary, the factors of final  $f_0$  movement and speech rate were adopted in a different way across six male native speakers depending on the politeness level, and were confirmed to have an actual impact on the perception of politeness. However, the impact of these two factors alone was not found to be powerful enough to override stylistic differences. (The stylistic differences here include the whole  $f_0$  contour, local tempo and loudness variations, articulation and voice quality.) This was especially evident in the attempt to make impolite utterances sound polite: the utterances of two speakers which were meant to be non-polite were judged neutral or very slightly polite with the 'polite' final prosody, and changes in speech rate could not substantially increase the politeness level of the original utterances. Indeed, changes in speech rate had the effect of making polite utterances less polite.

Since there are usually clear differences between polite utterances and non-polite utterances, which native speakers could easily report, there must be other

factors which are responsible for the stylistic differences, which have not as yet been identified. Further research is clearly needed. However, one of the problems which make this line of research very difficult is the effect of naturalness on politeness judgements. The results of Experiment 3 showed that naturalness judgements and politeness judgements tended to go hand in hand, especially when relatively minor effects caused by changes in certain acoustic variables were being examined, and cue manipulation using digital resynthesis often caused some kind of degradation and unnaturalness in the stimuli. The unnaturalness caused by such manipulations may create undesirable artefacts affecting subjects' perceptual judgements on the scales studied.

Finally, it is increasingly clear that listener characteristics must be carefully considered in politeness research. Thus, although there was a high level of inter-judge agreement on the scale of politeness in the present study, it was found that characteristics such as the speech rate of the subjects themselves had significant effects on their judgements. Consider one way in which this was manifest. The subjects showed a very high level of consistency in their naturalness responses, but there were clear subject differences. People appear to be very sensitive to unnaturalness by their own standards. Simply, this listener-specific sensitivity may bias politeness judgements. A single extreme value for any acoustic feature, (e. g., very fast rate of articulation), may reduce perceived politeness, but this will differ across listeners. This importance of perceived naturalness appears to influence the nature of politeness judgements: politeness judgements are likely to be made on confirming no negative features which would contribute to unnaturalness, rather than identifying positive features. In other words, polite utterances require that every influencing feature be kept within a certain range, which will vary from politeness level to politeness level and from speaker to speaker, and indeed from listener to listener.



## 7.2. Future work

In the present study very limited aspects of prosody were investigated in relation to politeness: final  $f_0$  movement in terms of duration and  $f_0$  direction, and global speech rate (which was realised by means of linear compression/expansion of segmental durations). The importance of the final  $f_0$  movement was confirmed. The effects of speech rate was also found to be significant, but contrary to my expectations, the rate factor was not powerful enough to override the style differences (i.e., 'polite' or 'casual' speaking style of the speaker). Slowing down utterances did not substantially contribute to higher perceived politeness scores. This was a slightly surprising result, because the speech rate factor was the most noticeable and consistent in both people's knowledge about politeness and production of polite speech. Slower speech was found to be the most noticeable feature for politeness in the large-scale questionnaire survey (Ogino and Hong, 1992), and all of our six speakers consistently adopted slower speech rates for polite utterances. So this less impressive impact of the speech rate variable may be due to the fact that the rate change in the experiments were crudely realised by linearly compressing or expanding each segmental duration, which almost never takes place in natural human speech (Section 3.5.1.3). Factors such as micro temporal structures (or local speech rate) and co-articulation would be relevant, and further experimentation is necessary.

The factor of speaking style, including local speech rate and intensity variation, articulation and voice quality, was not a focus of this study. However, the speaking style factor is apparently important for perceived politeness, as was suggested by the results of Experiment 2 and spectral analyses of speech samples, and thus deserves further investigation.

Other factors which are important in politeness research, but were excluded

from the present study are the factors of linguistic form, speech act of the utterance (e.g., request, apology) and the sex of the speaker. All the utterances used were simple question sentences spoken by male speakers. Among them, the interaction between paralinguistic factors and linguistic factors are particularly important. The linguistic factors include what kind of linguistic forms is used (e.g., honorific or plain) and whether or not softening expressions such as "erm...." and "if possible" are used in the utterance. In actual speech the verbal form and content of utterances (i.e., what you said) are inseparable from the tone of voice (i.e., how you said it), and hence indispensable in politeness judgement. Therefore, the effects of these linguistic factors and their interactions with paralinguistic factors will have to be addressed in future research, in order to fully understand the politeness judgement process.



## REFERENCES

- Abe, M., & Sato, H. (1993). Kotonaru hatsuwa youshiki niokeru inritsu tokuchou no bunseki [Prosodic characteristics of different speaking styles]. *Proc. Conference of the Acoustical Society of Japan*, March, 1993, 1-8-25 (pp. 193-194).
- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37, 715-727.
- Atal, B. S., & Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, 50, 637-655.
- Banse, R., & Scherer, K. R. (1996) Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614-636.
- Bell-Berti, F., Gelfer, C. E., & Boyle, M. (1995). Utterance-final lengthening: the effect of speaking rate. *Proc. ICPHS '95*, Stockholm, Vol. 1, pp. 162-165.
- Bezooijen, R. van (1984). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Dordrecht, The Netherlands: Foris.
- Bickley, C. (1972). Acoustic analysis and perception of breathy vowels. *MIT working papers in Speech Communication*, 1, 73-83.
- Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies: a cross-cultural study of speech act realization patterns. *Applied Linguistics*, 5(3), 196-212.
- Bolinger, D. (1964). Intonation as a universal. *Proc. The 9th International Congress of Linguists in 1962*, pp. 833-844.
- Borden, G. J., & Harris, K. S. (1984). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech* (2d ed.). Baltimore, MD: Williams & Wilkins.

- Brandt, J. F. (1972). Effects of stimulus bandwidth on listener judgments of vocal loudness and effect. *Journal of the Acoustical Society of America*, 52, 705-707 (Letter to the editor).
- Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse Intonation and Language Teaching*. London: Longman.
- Brody, M. W. (1943). Neurotic manifestations of the voice. *Psychoanalytic Quarterly*, 12, 371-380.
- Brown, B. L. (1980). Effects of speech rate on personality attributions and competency Evaluations. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social Psychological Perspectives* (pp. 293-300). Oxford: Pergamon Press.
- Brown, B. L., & Bradshaw, J. M. (1985). Towards a social psychology of voice variations. In H. Giles & R. N. St. Clair (Eds.), *Recent Advances in Language, Communication and Social Psychology* (pp. 144-181). London: Lawrence Erlbaum Associates.
- Brown, B. L., Giles, H., & Thakerar, J. N. (1985). Speaker evaluations as a function of speech rate, accent and context. *Language and Communication*, 5(3), 207-220.
- Brown, B. L., Strong, W. J., & Rencher, A. C. (1974). Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America*, 55, 313-318.
- Brown, P. (1980). How and why are women more polite: some evidence from a Mayan community. In S. McConnel-Ginet, R. Borker, & N. Furman (Eds.), *Women and Language in Literature and Society* (pp. 111-136). New York: Praeger.
- Brown, P., & C. Fraser. (1979). Speech as a marker of situation. In K. R. Scherer & H. Giles (Eds.), *Social Markers in Speech* (pp. 33-62). Cambridge, UK: Cambridge University Press.



- Brown, P., & Levinson, S. (1978). Universals in language usage: politeness phenomena. In E. N. Goody (Ed.), *Questions and Politeness: Strategies in social interaction* (pp. 56-289). Cambridge, UK: Cambridge University Press.
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.
- Campbell, W. N. (1992). Multi-level speech timing control. PhD dissertation, Department of Experimental Psychology, University of Sussex.
- Chang, T. N. C. (1958). Tones and intonation in the Chengtu dialect. *Phonetica*, 2, 60-84.
- Charpentier, F. J., & Stella, M. G. (1986). Diphone synthesis using an overlap-add technique for speech waveforms concatenation. *Proc. ICASSP '86*, Tokyo, pp. 2015-2018.
- Childers, D. G., & Lee, C. K. (1991). Vocal quality factors: analysis, synthesis, and perception. *Journal of the Acoustical Society of America*, 90, 2394-2410.
- Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology*, 9, 864-881.
- Coulmas, F. (1981). "Poison to your soul": thanks and apologies contrastively viewed. In F. Coulmas (Ed.), *Conversational routine* (pp. 69-91). The Hague: Mouton.
- Cruttenden, A. (1986). *Intonation*. Cambridge, UK: Cambridge University Press.
- Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge, UK: Cambridge University Press.
- Daniloff, R. G., & Hammarberg, R. (1974). On defining coarticulation. *Journal of Phonetics*, 1, 185-194.
- Davitz, J. R., & Davitz, L. J. (1959). The communication of feelings by content-free speech. *Journal of Communication*, 9, 6-13.

- Dusen, C. R. van (1941). A laboratory study of the metallic voice. *Journal of Speech Disorders*, 6, 137-140.
- Edelsky, C. (1979). Question intonation and sex roles. *Language in Society*, 8, 15-32.
- Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, 38, 270-277.
- Entropic Research Laboratory (1993). *Waves+/ESPS* (Ver. 5.0). Washington, DC: Entropic Research Laboratory.
- Fairbanks, G. (1940). Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America*, 11, 457-466.
- Fairbanks, G. M., & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monographs*, 6, 87-104.
- Fraser, B., & Nolan, W. (1981). The association of deference with linguistic form. *International Journal of the Sociology of Language*, 27, 98-111.
- Frick, R. W. (1985). Communicating emotion: the role of prosodic features. *Psychological Bulletin*, 97, 412-429.
- Frøkjær-Jensen, B., & Prytz, S. (1976). Registration of voice quality. *Brüel and Kjaer Technical Review*, 3, 3-17.
- Geluykens, R. (1987). Intonation and speech act type: an experimental approach to rising intonation in queclatives. *Journal of Pragmatics*, 11, 483-494.
- Geluykens, R., & Swerts, M. (1992). Prosodic topic- and turn-finality cues. *Proc. IRCS Workshop on Prosody in Natural Speech*, University of Pennsylvania, pp. 63-69.
- Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. New York: Anchor Books.



- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85, 845-857.
- Hart, J. 't, Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge, UK: Cambridge University Press.
- Henton, C. (1995). Pitch dynamism in female and male speech. *Language and Communication*, 15(1), 43-61.
- Henton, C. G., & Bladon, R. A. W. (1985). Breathiness in normal female speech: inefficiency versus desirability. *Language and Communication*, 5, 221-227.
- Hill, B., Ide, S., Ikuta, S., Kawasaki, A., & Ogino, T. (1986). Universals of linguistic politeness: quantitative evidence from Japanese and American English. *Journal of Pragmatics*, 10, 347-371.
- Holmes, J. N. (1985). A parallel-formant synthesizer for machine voice output. In F. Fallside & W. A. Woods (Eds.), *Computer Speech Processing* (pp. 163-187). London: Prentice-Hall International.
- Holmes, J. N. (1986). "SYNCON" Operating Instructions (Vers.3 and 4): A synthesis-by-rule software package for convenient interactive control of the Loughborough Sound Images Speech Synthesizer using a BBC microcomputer. Middlesex, UK: By the Author.
- Holmes, J. N., Mattingly, I. G., & Shearme, J. N. (1964). Speech synthesis by rule. *Language and Speech*, 7, 127-143.
- Hong, M. (1992). Kankoku-jin gakushuusha no nihongo no teinei hyougen ni mirareru inritsu-teki tokuchou [Prosodic characteristics of polite utterances in Japanese by Korean learners of Japanese], *Nihongo to nihon bungaku*, University of Tsukuba, 17, 32-42.
- Hong, M. (1993). Teinei hyougen ni okeru nihongo onsei no teineisa no kenkyuu [A study on Japanese phonetic politeness in polite expressions]. *Nihon onsei gakkaiishi*, No. 204, 13-30.

- Howell, D. C. (1992). *Statistical Methods for Psychology* (3d ed.). Belmont, CA: Duxbury Press.
- Ide, S. (1986). Introduction: the background of Japanese sociolinguistics. *Journal of Pragmatics*, 10, 281-286.
- Ide, S., Hill, B., Carnes, Y. M., Ogino, T., & Kawasaki, A. (1992). The concept of politeness: an empirical study of American English and Japanese. In R. Watts, S. Ide, & K. Enlich (Eds.), *Politeness in Language: Studies in its History, Theory and Practice* (pp. 281-297). Berlin: Mouton de Gruyter.
- Ide, S., Ogino, T., Kawasaki, A., & Ikuta, S. (1986). *Nihonjin to Amerikajin no keigo koudou* [politeness related behaviours of Japanese and Americans]. Tokyo: Nan'undou.
- Imaizumi, S., Hayashi, A., & Degushi, T. (1994). Listener-adaptive characteristics in dialogue: effects of temporal adjustments on emotional aspects of speech. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics*, No. 28, 59-64.
- Itakura, F., & Saito, S. (1968). Analysis-synthesis telephony based on the maximum likelihood method. *Proc. The 6th International Congress of Acoustics*, Tokyo, C-5-5.
- Kindaichi, H. (1964). Hanashi kotoba no keigo-teki hyougen [Honorific expressions in speech], *Gengo-seikatsu*, 149.
- Klasmeyer, G., & Sendlmeier, W. F. (1995). Objective voice parameters to characterize the emotional content in speech. *Proc. ICPHS '95*, Stockholm, Vol. 1, pp. 182-185.
- Klatt, D. H. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, 82, 737-793.
- Kramer, E. (1963). Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin*, 60, 408-420.



- Krom, G. de (1994). Spectral correlates of breathiness and roughness for different types of vowel fragments. *Proc. ICSLP'94, Yokohama, Vol. 3*, pp. 1471-1474.
- Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington, Indiana: Indiana University Press.
- Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type: voice quality and f<sub>0</sub> range in signaling speaker affect. *Journal of the Acoustical Society of America*, 78, 435-444.
- Lakoff, R. (1973). The logic of politeness; or, minding your p's & q's. *Papers from the 9th regional meeting of the Chicago Linguistic Society*, pp. 292-305.
- Lakoff, R. (1989). The limits of politeness: therapeutic and courtroom discourse. *Multilingua*, 8, 101-130.
- Laver, J. (1968). Voice quality and indexical information. *British Journal of Disorders of Communication*, 3, 43-54.
- Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge, UK: Cambridge University Press.
- Laver, J., & Hutcheson, S. (1972). Introduction. In J. Laver & S. Hutcheson (Eds.), *Communication in Face to Face Interaction* (pp. 11-15). Harmondsworth, UK: Penguin Books.
- Lawrence, W. (1953). The synthesis of speech from signals which have a low information rate. In W. Jackson (Ed.), *Communication Theory* (pp. 460-469). London: Butterworths.
- Leech, G. N. (1983). *Principles of Pragmatics*. London: Longman.
- Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America*, 34, 922-927.

- Loveday, L. (1981). Pitch, politeness and sexual role: an exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, 24, 71-89.
- LSI (1990). *LSI Speech Workstation* (Ver. 1.0). Loughborough, Leices., UK: Loughborough Sound Images, Ltd.
- Mallory, E., & Miller, V. A. (1958). A possible basis for the association of voice characteristics and personality traits. *Speech Monographs*, 25, 255-260.
- Mao, L. R. (1994). Beyond politeness theory: 'face' revised and renewed. *Journal of Pragmatics*, 21, 451-486.
- Martin, S. E. (1964). Speech levels in Japan and Korea. In D. Hymes (Ed.), *Language in Culture and Society* (pp. 407-414). New York: Harper and Row.
- Martin, S. E. (1975). *A reference grammar of Japanese*. New Haven, CT: Yale University Press.
- Matsumoto, Y. (1988). Reexamination of the universality of face: politeness phenomena in Japanese. *Journal of Pragmatics*, 12, 403-426.
- McLemore, C. A. (1992). Prosodic variation across discourse types. *Proc. IRCS Workshop on Prosody in Natural Speech*, University of Pennsylvania, pp. 117-128.
- Menn, L., & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, 25, 341-383.
- Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica*, 41, 215-225.
- Minami, F. (1987). *Keigo* [Honorific expressions]. Tokyo: Iwanami Shinsho.
- Miyaji, H. (1985). Taiguu hyougen [Expressions of addressing people]. In NLRI [Kokuritsu Kokugo Kenkyuujo] (Ed.), *Nihongo kyoushi-you sankousho 1: Gengo koudou to nihongo kyouiku*. Tokyo: Bonjinsha.



- Miyatake, M., & Sagisaka, Y. (1990). Shuju no hatsuwa youshiki ni mirareru inritsu tokuchouto sono seigyō [Prosodic characteristics and their control in Japanese speech with various speaking styles], *Nihon denshi jouhou tsuushin gakkai ronbunshi*, Vol. J73-D-II, No. 12, pp. 1929-1935.
- Monsen, R. B., & Engebretson, A. M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America*, 62, 981-993.
- Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9, 453-467.
- Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93, 1097-1108.
- Murray, I. R., Arnott, J. L., & Newell, A. F. (1988). HAMLET - simulating emotion in synthetic speech. *Proc. Speech'88, The 7th FASE Symposium*, Edinburgh, Vol. 4, pp. 1217-1223.
- Nakane, C. (1967). *Tate-shakai no ningen kankei: tan'itsu shakai no riron* [Personal relations in a vertical society: a theory of a homogeneous society]. Tokyo: Koudansha.
- Nakane, C. (1970). *Japanese Society*. Berkeley, CA: University of California Press.
- NHK (Ed.). (1995). *NHK anaunsaa no sutekina hanashi kotoba* [NHK announcers' nicely spoken language]. Tokyo: NHK Press.
- NLRI [Kokuritsu Kokugo Kenkyuujo] (Ed.) (1957). *Keigo to keigo ishiki* [Honorific expressions and people's knowledge about the honorific system] (NLRI Report 11). Tokyo: Shuuei-shuppan.
- NLRI [Kokuritsu Kokugo Kenkyuujo] (Ed.) (1971). *Taiguu hyougen no jittai - Matsue 24 jikan chousa shiryō kara* - [Usage of the honorific system: from 24-hour recordings in Matsue] (NLRI Report 41). Tokyo: Shuuei-shuppan.

- NLRI [Kokuritsu Kokugo Kenkyuujo] (Ed.) (1982). *Kigyō no nakano keigo* [Honorific expressions in a workplace] (NLRI Report 73). Tokyo: Sanseidou.
- NLRI [Kokuritsu Kokugo Kenkyuujo] (Ed.) (1983). *Keigo to keigo ishiki - Okazaki ni okeru 20-nenmae tonō hikaku* - [Honorific expressions and people's knowledge about the honorific system: a comparison with the data of Okazaki survey 20 years ago] (NLRI Report 77). Tokyo: Sanseidou.
- NLRI [Kokuritsu Kokugo Kenkyuujo] (Ed.) (1986). *Shakai henka to keigo koudou no hyōjun* [Changes in society and the standard of politeness related behaviours] (NLRI Report 86). Tokyo: Shuei-shuppan.
- Nomoto, K. (1974). *Keigo no kenkyū - Chōsa · Bunseki no hōhō* [Research on honorific systems - methods of survey and analysis]. In *Keigokouza 10: Keigo kenkyū no hōhō*. Tokyo: Meiji-shoin.
- Ofuka, E., Valbret, H., Waterman, M., Campbell, N., & Roach, P. (1994). The role of f<sub>0</sub> and duration in signalling affect in Japanese: anger, kindness and politeness. *Proc. ICSLP'94, Yokohama, Vol. 3*, pp. 1447-1450.
- Ogino, T., & Hong, M. (1992). *Nihongo onsei no teineisa ni kansuru kenkyū* [A study of politeness in Japanese speech]. In T. Kunihiro (Ed.), *Nihongo intonation no jittai to bunseki* [The state-of-the-art, and analysis, of Japanese intonation] (pp. 215-258). Tokyo: Monbushō.
- Ohala, J. J. (1984). An ethological perspective on common cross-language utilisation of F<sub>0</sub> of voice. *Phonetica*, 41, 1-16.
- Otsubo, K. (1990). *Onsei kyouiku no mondaiten* [Problems of spoken language education]. In *Kouza nihongo to nihongo kyouiku 3: Nihongo no onsei · on'in (ge)*. Tokyo: Meiji-shoin.
- Pittam, J., & Scherer, K. R. (1993). Vocal expression and communication of emotion. In M. Lewis & J. Haviland (Eds.), *The handbook of emotions* (pp. 185-198). New York: Guilford.
- Rosen, G. (1958). A dynamic analog speech synthesizer. *Journal of the Acoustical Society of America*, 30, 201-209.



- Rosenthal, R. (1982). Conducting judgment studies. In K. R. Scherer & P. Ekman (Eds.), *Handbook of method in nonverbal behavior research* (pp. 287-361). Cambridge, UK: Cambridge University Press.
- Ross, E. D., Edmondson, J. A., & Seibert, G. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: a comparison based on computer analysis of voice. *Journal of Phonetics*, 14, 283-302.
- Scherer, K. R. (1979a). Nonlinguistic vocal indicators of emotion and psychopathology. In C. E. Izard (Ed.), *Emotions in Personality and Psychopathology* (pp. 495-529). New York: Plenum Press.
- Scherer, K. R. (1979b). Personality markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social Markers in Speech* (pp. 147-209). Cambridge, UK: Cambridge University Press.
- Scherer, K. R. (1982). Methods of research on vocal communication: paradigms and parameters. In K. R. Scherer & P. Ekman (Eds.), *Handbook of Methods in Nonverbal Behavior Research* (pp. 136-198). Cambridge, UK: Cambridge University Press.
- Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: testing two models. *Journal of the Acoustical Society of America*, 76, 1346-1356.
- Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, 1(4), 331-346.
- Scherer, U., & Scherer, K. R. (1980). Psychological factors in bureaucratic encounters: determinants and effects of interactions between officials and clients. In W. T. Singleton, P. Spurgeon, & R. B. Stammers (Eds.), *The Analysis of Social Skill* (pp. 315-328). New York: Plenum Press.
- Shibata, T., Ogino, T., Fujita, K., Ozaki, I., & Misonou, Y. (1980). Toshi no keigo no shakai gengo-teki kenkyuu - Shouwa 53-nendo Sapporo ni okeru keigo chousa houkoku - [Sociolinguistic research on the Japanese honorific system in cities - Report on the survey in Sapporo in 1978]. In Monbushou Tokutei

Kenkyuu "gengo" soukatsu-han (Ed.), *"Gengo" Kenkyuu seika kankousho*. Tokyo: Monbushou.

Skinner, E. R. (1935). A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness. *Speech Monographs*, 2, 81-137.

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech*, 18, 145-152.

Starkweather, J. A. (1956). Content-free speech as a source of information about the speaker. *Journal of Abnormal and Social Psychology*, 52(3), 394-402.

Stevens, K. N., Kasowski, S., & Fant, G. (1953). An electric analog of the vocal tract. *Journal of the Acoustical Society of America*, 25, 734-742.

Stross, B. (1977). Tzeltal conceptions of power. In R. D. Fogelson & R. N. Adams (Eds.), *The anthropology of power: Ethnographic studies from Asia, Oceania, and the New World* (pp. 271-285). New York: Academic Press.

Sugito, S. (1981). Aisatsu no kotoba to miburi [Expressions and gestures in greetings]. In Bunkachou (Ed.), *Aisatsu to kotoba* (Kotoba series 14). Tokyo: Bunkachou.

Sugitou, M. (1986). Nyuusu no houdouni okeru hatsuwa jikan oyobi kyuushi jikan to hatsuwa sokudo [The relationship between speech/pause time and speech rate in news broadcast]. Tokyo: Shoin Kokubungaku.

Takefuta, Y. (1975). Method of acoustic analysis of intonation. In S. Singh (Ed.), *Measurement Procedures in Speech, Hearing and Language* (pp. 363-378). Baltimore, MD: University Park Press.

Terango, L. (1966). Pitch and duration characteristics of the oral reading of males on a masculinity-femininity dimension. *Journal of Speech and Hearing Research*, 9, 590-595.

Uldall, E. (1964). Dimensions of meaning in intonation. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In Honour of*



*Daniel Jones: Papers Contributed on the Occasion of his Eightieth Birthday, 12 September 1961* (pp. 271-279). London: Longman.

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 51(4), 690-699.

Watkins, A. J., & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *Journal of the Acoustical Society of America*, 96, 1263-1282.

Watts, R. J., Ide, S., & Enlich, K. (1992) Introduction. In R. J. Watts, S. Ide, & K. Enlich (Eds.), *Politeness in Language: Studies in its History, Theory and Practice* (pp. 1-17). Berlin: Mouton de Gruyter.

Wendahl, R. W. (1963). Laryngeal analog synthesis of harsh voice quality. *Folia Phoniatica*, 15, 241-250.

Wendahl, R. W. (1964). The role of amplitude breaks in the perception of vocal roughness. *American Speech and Hearing Association Convention Abstracts*, 6, 406.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, 52, 1238-1250.

# APPENDIX 1

## **A pilot study: Application to synthesis**



## Aim

A small-scale listening test was conducted, in order to examine people's responses to prosodic features in synthetic speech politeness judgements.

## Method

### *1. Speech material and stimulus preparation*

There were 13 different stimulus conditions of a single sentence produced with an adult male voice, using the synthesis-by-rule software package SYNCON on a BBC microcomputer (Holmes, 1986; Holmes *et al.*, 1964; Holmes, 1985). The prosodic factors varied were (a) segmental duration, (b)  $f_0$  contour and (c) intensity.

The sentence used is 'Kokokara Ginza made(,) donokurai kakarun deshouka' meaning 'from here to Ginza, how long would it take?'. This sentence ('Ginza' sentence) was selected because first, longer utterances were thought to be better to examine the effects of prosody when the utterances were not very natural, and the 'Ginza' sentence is reasonably long (consisting of 22 morae); second, politeness is expected to be important and meaningful for this sentence in the sense that the reply may depend on how the question is asked; finally, the 'Ginza' sentence was used in Ogino and Hong's (1992) study and their recordings of the polite and non-polite utterances of the sentence spoken by 12 speakers, with perceived politeness scores rated by 200 or so Japanese people, were available.

The SYNCON synthesiser, which uses the formant synthesis technique (Section 3.5.1.1.2), allows users to change various parameters including sound element (i.e., vowel and consonant) qualities, segmental duration,  $f_0$  and intensity. All parameter values can be updated every 10 ms, and the system has default values for all

parameters including those for determining the sound element quality for English. These default values for the English sound elements were used for reconstructing the Japanese utterance. Although sounds used in Japanese are slightly different from those for English, the output utterances were perfectly intelligible. However, the utterances were not quite natural mainly because of slightly mechanical articulation or/and speech quality, sounding like a robot's speaking Japanese. The duration variable can be changed in 10 ms steps. The  $f_0$  parameter has values in the range between 1 to 63, which control  $f_0$  values on a logarithmic scale from 27.3 Hz to 400 Hz. The modification of intensity level was made through a built-in interface. Since there was little evidence that intensity was a significant cue for signalling affect (Section 3.3.3.3), this variable was included only to examine how people would respond to differences in intensity. The values for this variable could range from -18 dB to +18 dB in 2 dB steps. One step change increases/decreases all amplitude levels of the element by 2 dB, except the low-frequency amplitude (below the first formant), which is changed by 1 dB.

The segmental durations and  $f_0$  values of the polite and non-polite utterances spoken by a trained female speaker were measured using the digital signal processing package LSI (LSI, 1990), and were used as 'polite'/'non-polite' prosody. In spite of a male voice used for synthesis, a female speaker's utterances were used, because this speaker differentiated her polite and non-polite versions very clearly: her polite version was rated as the second politest (Mean politeness score = 2.63) and her non-polite one, the least polite (Mean politeness score = 1.09) on a politeness scale ranging from 1 (not polite at all) to 4 (very polite), among 12 speakers in Ogino and Hong's (1992) study. However, there was some concern that using prosody of female utterances for male voices might introduce an artefact because of some differences in use of certain aspects of prosody (e.g.,  $f_0$  range) between female and male speakers (as was seen in Section 3.3.2). In spite of this concern, it was judged that the effects of using female



prosody for a 'male' voice would be minor and using clearly different prosody would be more beneficial due to the fact that the synthetic utterances used did not sound like natural human speech. The waveforms and  $f_0$  contours of the original polite and non-polite utterances are attached in Appendix J.

#### (a) Segmental duration

Four types of segmental duration including duration of pause were used: (1) D(P): durations taken from the polite utterance of the female speaker, (2) D(I): durations from the non-polite utterance of the same speaker, (3) 0.8D(P): 20% linearly compressed durations of the D(P) type and (4) 0.7D(P): 30% linearly compressed durations of the D(P) type. The original polite version had a long pause about 260 ms between Phrase 1 ('kokokara ... made') and Phrase 2 ('Ginza ... deshouka') while the original non-polite version had little pause (less than 50 ms). The speech rate (exclusive of pause and final morae of the two phrases) of the polite version was 10 mora per sec, and that of the non-polite version was 12 mora per sec.

#### (b) F0 contour

Four types of the  $f_0$  contour were used: (1) F0(P):  $f_0$  contour of the polite utterance, (2) F0(I):  $f_0$  contour of the non-polite utterance, (3) F0(P/): the F0(P) type  $f_0$  contour with its  $f_0$  movement of the final vowel being replaced with the F0(I) type movement and (4) F0(I\): the F0(I) type contour with its final  $f_0$  movement being replaced with the F0(P) type movement. The final  $f_0$  direction of the polite utterance was a fall (which is a default tone for the expression '-deshouka'), and that of the non-polite utterance was a rise. All the  $f_0$  values taken from the original female utterances were then lowered slightly (by about 0.6 octave) to the male register, because a reasonably natural-sounding female speech could not be achieved by the synthesiser.

The mean  $f_0$  of  $F_0(P)$  was 150 Hz and that of  $F_0(I)$  was 160 Hz. The  $F_0(P/)$  type was only realised with such duration types as  $D(P)$  and  $0.7D(P)$ , and the  $F_0(I/)$  type, with the  $D(I)$  type.

### (c) Intensity

Three types of intensity values were used: (1) the synthesiser's default values for the elements, (2) \*P type and (3) \*I type. Since there was no direct way of realising the real loudness, default intensity values were modified so that the intensity value of each phoneme was roughly proportional to the rms-energy level of the original utterances. The \*P type modelled on the polite utterance and the \*I type, on the non-polite utterance. The alteration of intensity resulted in slightly increasing the average level of intensity of the utterances of both types: the mean rms-energy across voiced segments of the \*P type was 1.3 times greater, and that of \*I type was 1.7 times greater than the energy level of utterances with the default intensity values. The intensity alteration was only applied to the stimulus conditions  $D(P)F_0(P)$  and  $D(I)F_0(I)$ .

These three prosodic features mentioned above are summarised in Table 1, and the waveforms and  $f_0$  contours of  $D(P)F_0(P)$  and  $D(I)F_0(I)$  are shown in Appendix J. A total of 14 stimuli, consisting of 13 stimulus conditions and a dummy stimulus, which was located at the beginning, were recorded only once on tape in random order. Each item was followed by a 6-second silence during which subjects were asked to make politeness judgements. Only one occurrence for each condition was used, because it was considered to be desirable to obtain responses which were close to the first impression. It was expected that familiarity with the synthetic speech could influence people's responses substantially. In other words, the longer subjects are exposed to synthetic speech, the more natural the speech becomes, and this change in naturalness could affect their politeness judgement system.



TABLE 1. Prosodic characteristics of the 13 stimulus conditions in SYNCON experiment.

<i>Condition</i>		<i>Duration</i>		<i>F0</i>		<i>Intensity</i>
<i>Duration type</i>	<i>F0 type</i>	<i>Speech rate*<sup>1</sup> (mora/s)</i>	<i>Pause (ms)</i>	<i>Contour type</i>	<i>Final direction</i>	
D(P)	F0(P)	9.9	260	P	fall	-
	F0(P)*			P	fall	*P type
	F0(P/)			P	rise	-
	F0(I)			I	rise	-
D(I)	F0(I)	12.2	30	I	rise	-
	F0(I)*			I	rise	*I type
	F0(I/)			I	fall	-
	F0(P)			P	fall	-
0.8D(P)	F0(P)	12.4	210	P	fall	-
	F0(I)			I	rise	-
0.7D(P)	F0(P)	14.1	180	P	fall	-
	F0(P/)			P	rise	-
	F0(I)			I	rise	-

\*1: Pauses and final morae in Phrase 1 and Phrase 2 of the sentence (i.e., 'de' and 'ka') were excluded from speech rate calculation.

## 2. Rating scales

The bipolar 5-point scales shown in Fig. 1 were used for politeness, tempo and naturalness ratings. The scales of tempo and naturalness were included to assess whether or not the stimuli were perceived as intolerably unnatural to the subjects in terms of both tempo and speech quality.

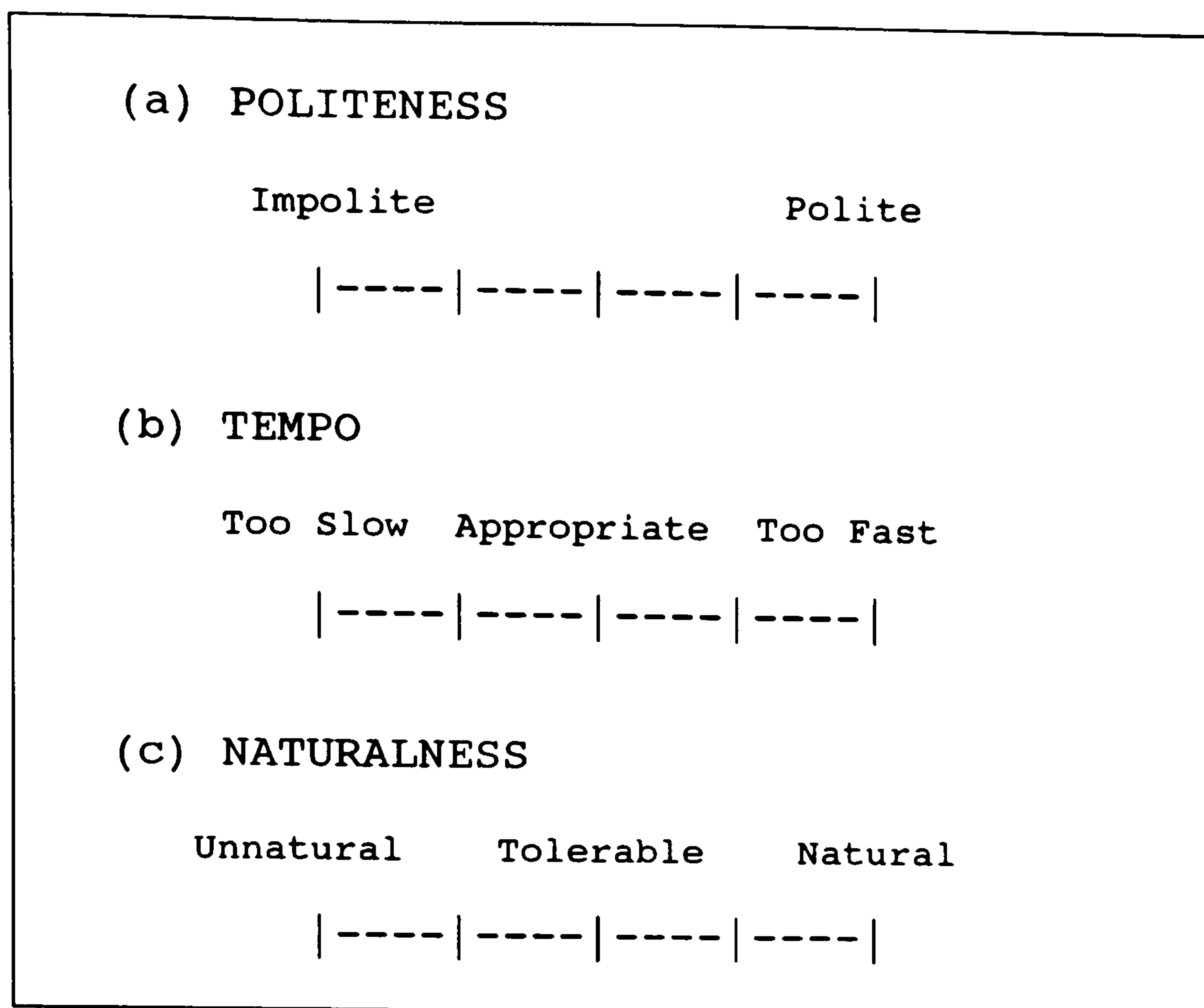


FIG. 1. Rating scales for politeness (a), tempo (b) and naturalness (c).

### 3. Subjects

Five female subjects, ranging in age between 37 and 62, participated in the listening test voluntarily. Female subjects who were in the relatively older age range were selected, because females had been known to be better encoders of non-verbal cues than males (e.g., Hall, 1978) and older generations were generally more polite in speech. All the subjects were well educated native speakers of Japanese. Three were from the Tokyo area. One subject (F1) was born in Tokyo, but living in the western area for more than 20 years, and one (F5) was originally from the northern part of Japan, but working in Tokyo. None of the subjects was familiar with synthetic speech.

### 4. Rating sessions

Subjects were given instructions (i.e., what they were supposed to do in the test) orally at the beginning of the session. They were told that the speaker was asking a person in the street this 'Ginza' question. Although the first impressions were desirable,



it was considered to be very difficult to judge politeness on a scale with rather unexpected synthetic utterances. Therefore the tape was played once to subjects through headphones prior to the rating sessions, in order for them to become reasonably accustomed to the sounds. All subjects were presented with the stimuli in the same order. The five subjects were tested individually in a quiet room, and each session lasted about five minutes. After the politeness ratings, two of the subjects (F1 and F2) listened to the same tape again after a long break, and rated each utterance on the scales of tempo and naturalness during the 6-second silence following the utterance.

## **Results and discussion**

The rating scores for politeness, tempo and naturalness for each stimulus condition are shown in Table 2. The scores could range from -2 (very impolite/too slow/very unnatural) to +2 (very polite/too fast/very natural).

TABLE 2. Rating scores for politeness, tempo and naturalness for each condition.

Condition			PO *1	Politeness					Tempo		Natural- ness	
Type				Subjects (age)					Subjects		Subjects	
DUR	F0	INT		F1 (62)	F2 (55)	F3 (50)	F4 (37)	F5 (37)	F1	F2	F1	F2
P	P	*P	9	0	+1	+2	+1	0	-1	-1	0	+1
	P		10	0	+1	0	+1	0	0	-1	+0.5	0
	P/		8	+1	+1	+1	0	0	-1	-1	0	+1
	I		1	+1	0	0	0	0	-1	-1	0	0
I	I	*I	7	+1	0	-1	-1	0	0	0	+1	+1
	I		2	-1	-1	-1	-1	0	+2	+1	-1	-1
	I\		3	0	0	-1	+1	0	+1	+1	0	0
	P		5	0	0	-1	+1	0	0	0	0	+1
0.8P	P		0/13	0	0	+1	-1	0	0	0	+1	+1
	I		12	+1	0	0	-1	0	0	0	+1	+1
0.7P	P		11	-1	-1	-2	+1	0	+1.5	+1	-1	-1
	P/		6	+1	-1	-1	-1	0	+0.5	+1	+0.5	0
	I		4	+1	0	-2	-1	0	+1	+1	0	0

\*1: PO means presentation order of the stimuli. PO = 0 means the stimulus was a dummy.

DUR: duration type, F0: f0 type and INT: intensity type.

(a) Politeness ratings

Great variability was found among the five subjects' ratings. F5 did not distinguish any conditions, whereas the other four subjects did, but in different ways. Table 3 (a) shows politeness scores by speech rate and f0 contour types. F2 and F3 responded to the speech rate, while F1 and F4 rather responded to the contour types. F1 rated a final rise more positively while F4 preferred a final fall (Table 3 (b)).

In order to assess the effects of speech rate, scores of F2 and F3 were examined. Table 3 (a) shows that the slower the speech rate is, the more polite the utterance was



rated: the D(P) type (Speech rate: 10 mora per sec) was rated as the most polite and the 0.7D(P) type (Speech rate: 14 mora per sec) the most impolite. The effects of the micro temporal structure on politeness ratings was examined by comparing the scores for the D(I) type and 0.8D(P) type, the speech rate of both of which was roughly the same (about 12 mora per sec) (Table 3 (c)). Although F2 did not distinguish these two conditions, F3 rated the 0.8D(P) type more positively. The main difference between these duration conditions was the presence or absence of pause: the D(I) type had no noticeable pause while the 0.8D(P) type had a 210-ms pause. So pausing appears to have effects on politeness judgements to some people. The acoustic analyses (Chapter 5) also confirmed that speakers tended to insert a pause when they intended to speak politely.

The conditions in which the intensity values were slightly increased (i.e., D(P)F0(P)\* and D(I)F0(I)\*) were rated as either the same or less polite. The naturalness scores by the two subjects show that the \*P type sounded reasonably natural, but the \*I type was perceived as very unnatural. It appears to suggest that increasing loudness makes speech less polite, or will exaggerate the impoliteness if the utterance was already impolite.



TABLE 3. Politeness scores by speech rate, f0 contour type and f0 final direction.

(a) Politeness scores by speech rate and f0 contour type, rated by four subjects (F1 ~ F4).

<i>Condition</i>	<i>Speech rate (mora/sec)</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>
D(P)F0(P)	9.9	0	+1	+2	+1
D(P)F0(I)	9.9	+1	0	0	0
0.8D(P)F0(P)	12.4	0	0	+1	-1
0.8D(P)F0(I)	12.4	+1	0	0	-1
0.7D(P)F0(P)	14.1	-1	-1	-2	+1
0.7D(P)F0(I)	14.1	+1	0	-2	-1

(b) Politeness scores by final f0 direction type, rated by four subjects (F1 ~ F4).

<i>Condition</i>	<i>final direction</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>
D(P)F0(P)	fall	0	+1	+2	+1
D(P)F0(P/)	rise	+1	+1	+1	0
D(I)F0(I/)	fall	0	0	-1	+1
D(I)F0(I)	rise	+1	0	-1	-1
0.7D(P)F0(P)	fall	-1	-1	-2	+1
0.7D(P)F0(P/)	rise	+1	0	-2	-1

(c) Comparison between the politeness ratings for D(I) and those for 0.8D(P).

<i>Condition</i>	<i>F2</i>	<i>F3</i>
D(I)F0(P)	0	-1
D(I)F0(I)	0	-1
0.8D(P)F0(P)	0	+1
0.8D(P)F0(I)	0	0



### (b) Tempo and naturalness ratings

Although the politeness ratings of F1 and F2 were different, the ratings of tempo and naturalness of both subjects were quite consistent. All the conditions except the condition of increased intensity (the \*I type) and one of the conditions of the fastest rate (0.7D(P), speech rate: 14 mora/sec) were rated tolerable or even natural. The reason why only one condition of the fastest rate was rated unnatural may be because of the contrast effect; the previous stimulus to this condition was the slowest one (speech rate: 10 mora/sec), while the previous stimuli to the other fastest conditions were the D(I) type (speech rate: 12 mora/sec). The tempo ratings agree with the actual speech rate, with the 0.8D(P) type and D(I) being rated as most appropriate and also natural. However, this appropriateness in tempo and naturalness does not appear to lead high politeness scores directly: F2, who responded to speech rate in her politeness ratings, rated this condition as neutral while she rated most of the slower versions as polite.

### **Conclusion**

Although some caution is needed to interpret the results obtained by this listening test due to the small number of subjects here, the results show that most subjects were sensitive to certain aspects of prosody and did judge politeness using prosodic features, even if the utterances were not as natural as normal human speech. The effects of prosody can be substantial for any kind of utterance (e.g., either synthetic or foreign) in signalling politeness. The judgement criteria, however, appear to vary from person to person. In this experiment there were three types of responses to three prosodic factors studied (i.e., duration, f0 contour and intensity): (1) responding to mainly duration, (2) responding to mainly f0 contour, especially final f0 direction, and (3) not responding to prosody. This variability, however, might be explained more elegantly

by introducing the linguistic/social characteristics of subjects (e.g., age, accent).

Further experimentation with more subjects is needed to clarify this point.



# APPENDIX 2

## ANOVA results

**(1-A-1) ANOVA results in Experiment 1-A:**

**Speaker, Style, Final Duration(F\_DUR) and Final F0 Direction(F\_F0)  
as within-subjects factors**

SPSS/PC+ The Statistical Package for IBM PC

6/10/96

SET WORKDEV=c.

SET WKSPACE=1500.

GET /FILE 'eff01.dat'.

The SPSS/PC+ system file is read from  
file eff01.dat

The file was created on 7/17/95 at 14:06:13

and is titled SPSS/PC+ System File Written by Data Entry II

The SPSS/PC+ system file contains

20 cases, each consisting of

150 variables (including system variables).

150 variables will be used in this session.

-----

MANOVA ks1 ks2 ks3 ks4 ks5 ks6 ks7 ks8 tk1 tk2 tk3 tk4 tk5 tk6 tk7 tk8  
/wsfactors speaker(2) style(2) f\_dur(2) f\_f0(2) /wsdesign.

20 cases accepted.

0 cases rejected because of out-of-range factor values.

0 cases rejected because of missing data.

1 non-empty cells.

0 design will be processed.

#### Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	144.23	19	7.59		
CONSTANT	5393.45	1	5393.45	710.50	.000

- - - - -

#### Tests involving 'SPEAKER' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	122.48	19	6.45		
SPEAKER	48.21	1	48.21	7.48	.013

- - - - -

#### Tests involving 'STYLE' Within-Subject Effect.

Tests of Significance for T3 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	140.35	19	7.39		
STYLE	5.92	1	5.92	.80	.382

- - - - -



## Tests involving 'F\_DUR' Within-Subject Effect.

Tests of Significance for T4 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	36.32	19	1.91		
F_DUR	66.40	1	66.40	34.73	.000

- - - - -

## Tests involving 'F\_F0' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	13.91	19	.73		
F_F0	6.87	1	6.87	9.38	.006

- - - - -

## Tests involving 'SPEAKER BY STYLE' Within-Subject Effect.

Tests of Significance for T6 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	33.27	19	1.75		
SPEAKER BY STYLE	19.30	1	19.30	11.02	.004

- - - - -

## Tests involving 'SPEAKER BY F\_DUR' Within-Subject Effect.

Tests of Significance for T7 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	15.70	19	.83		
SPEAKER BY F_DUR	8.44	1	8.44	10.22	.005

- - - - -

## Tests involving 'SPEAKER BY F\_F0' Within-Subject Effect.

Tests of Significance for T8 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	7.48	19	.39		
SPEAKER BY F_F0	.88	1	.88	2.23	.152

- - - - -

## Tests involving 'STYLE BY F\_DUR' Within-Subject Effect.

Tests of Significance for T9 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	6.09	19	.32		
STYLE BY F_DUR	8.18	1	8.18	25.54	.000

- - - - -

## Tests involving 'STYLE BY F\_F0' Within-Subject Effect.

Tests of Significance for T10 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
---------------------	----	----	----	---	----------

WITHIN CELLS	4.41	19	.23		
STYLE BY F_F0	.58	1	.58	2.52	.129

- - - - -

Tests involving 'F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T11 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	6.80	19	.36		
F_DUR BY F_F0	.00	1	.00	.00	.953

- - - - -

Tests involving 'SPEAKER BY STYLE BY F\_DUR' Within-Subject Effect.

Tests of Significance for T12 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	12.15	19	.64		
SPEAKER BY STYLE BY F_DUR	.84	1	.84	1.32	.265

- - - - -

Tests involving 'SPEAKER BY STYLE BY F\_F0' Within-Subject Effect.

Tests of Significance for T13 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	8.68	19	.46		
SPEAKER BY STYLE BY F_F0	.08	1	.08	.17	.688

- - - - -

Tests involving 'SPEAKER BY F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T14 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.87	19	.31		
SPEAKER BY F_DUR BY F_F0	.00	1	.00	.01	.909

- - - - -

Tests involving 'STYLE BY F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T15 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	3.22	19	.17		
STYLE BY F_DUR BY F_F0	.45	1	.45	2.64	.121

- - - - -

Tests involving 'SPEAKER BY STYLE BY F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T16 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F



WITHIN CELLS	7.59	19	.40		
SPEAKER BY STYLE BY	.64	1	.64	1.61	.219
F_DUR BY F_F0					
- - - - -					

(1-A-2) ANOVA results in Experiment 1-A:  
Speaker, Style, Final Duration(F\_DUR) and Final F0 Direction(F\_F0)  
as within-subjects factors, and  
Accent of Listener (AREA: Easter, Western and Others)  
as a between-subjects factor

MANOVA ks1 ks2 ks3 ks4 ks5 ks6 ks7 ks8 tk1 tk2 tk3 tk4 tk5 tk6 tk7 tk8  
by area(1,3) /wsfactors speaker(2) style(2) f\_dur(2) f\_f0(2)  
/wsdesign.

20 cases accepted.  
0 cases rejected because of out-of-range factor values.  
0 cases rejected because of missing data.  
3 non-empty cells.  
  
0 design will be processed.

- - - - -  
Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	70.45	17	4.14		
CONSTANT	3822.51	1	3822.51	922.41	.000
AREA	73.78	2	36.89	8.90	.002

- - - - -  
Tests involving 'SPEAKER' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	88.15	17	5.19		
SPEAKER	3.02	1	3.02	.58	.456
AREA BY SPEAKER	34.33	2	17.16	3.31	.061

- - - - -  
Tests involving 'STYLE' Within-Subject Effect.

Tests of Significance for T3 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	93.64	17	5.51		
STYLE	2.67	1	2.67	.48	.496
AREA BY STYLE	46.71	2	23.35	4.24	.032

- - - - -  
Tests involving 'F\_DUR' Within-Subject Effect.

Tests of Significance for T4 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	28.92	17	1.70		
F_DUR	21.25	1	21.25	12.49	.003
AREA BY F_DUR	7.40	2	3.70	2.18	.144

- - - - -

Tests involving 'F\_F0' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	12.82	17	.75		
F_F0	3.09	1	3.09	4.09	.059
AREA BY F_F0	1.09	2	.54	.72	.500

- - - - -

Tests involving 'SPEAKER BY STYLE' Within-Subject Effect.

Tests of Significance for T6 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	31.43	17	1.85		
SPEAKER BY STYLE	14.79	1	14.79	8.00	.012
AREA BY SPEAKER BY S TYLE	1.84	2	.92	.50	.616

- - - - -

Tests involving 'SPEAKER BY F\_DUR' Within-Subject Effect.

Tests of Significance for T7 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	11.86	17	.70		
SPEAKER BY F_DUR	2.47	1	2.47	3.54	.077
AREA BY SPEAKER BY F _DUR	3.83	2	1.92	2.75	.093

- - - - -

Tests involving 'SPEAKER BY F\_F0' Within-Subject Effect.

Tests of Significance for T8 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	6.49	17	.38		
SPEAKER BY F_F0	.20	1	.20	.52	.481
AREA BY SPEAKER BY F _FO	.99	2	.49	1.29	.301

- - - - -

Tests involving 'STYLE BY F\_DUR' Within-Subject Effect.

Tests of Significance for T9 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.68	17	.33		
STYLE BY F_DUR	3.30	1	3.30	9.87	.006
AREA BY STYLE BY F_D	.41	2	.20	.61	.556



UR

- - - - -

Tests involving 'STYLE BY F\_F0' Within-Subject Effect.

Tests of Significance for T10 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	4.23	17	.25		
STYLE BY F_F0	.42	1	.42	1.67	.213
AREA BY STYLE BY F_F0	.17	2	.09	.35	.710

- - - - -

Tests involving 'F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T11 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	6.71	17	.39		
F_DUR BY F_F0	.01	1	.01	.02	.881
AREA BY F_DUR BY F_F0	.09	2	.04	.11	.894

- - - - -

Tests involving 'SPEAKER BY STYLE BY F\_DUR' Within-Subject Effect.

Tests of Significance for T12 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	11.96	17	.70		
SPEAKER BY STYLE BY F_DUR	.57	1	.57	.81	.381
AREA BY SPEAKER BY STYLE BY F_DUR	.19	2	.10	.14	.873

- - - - -

Tests involving 'SPEAKER BY STYLE BY F\_F0' Within-Subject Effect.

Tests of Significance for T13 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	8.38	17	.49		
SPEAKER BY STYLE BY F_F0	.21	1	.21	.42	.526
AREA BY SPEAKER BY STYLE BY F_F0	.29	2	.15	.30	.748

- - - - -

Tests involving 'SPEAKER BY F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T14 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.17	17	.30		
SPEAKER BY F_DUR BY F_F0	.07	1	.07	.21	.649
AREA BY SPEAKER BY F_DUR BY F_F0	.70	2	.35	1.16	.338

- - - - -

Tests involving 'STYLE BY F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T15 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	2.77	17	.16		
STYLE BY F_DUR BY F_F0	.12	1	.12	.76	.396
AREA BY STYLE BY F_DUR BY F_F0	.45	2	.23	1.40	.275

- - - - -

Tests involving 'SPEAKER BY STYLE BY F\_DUR BY F\_F0' Within-Subject Effect.

Tests of Significance for T16 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	7.14	17	.42		
SPEAKER BY STYLE BY F_DUR BY F_F0	.41	1	.41	.98	.336
AREA BY SPEAKER BY STYLE BY F_DUR BY F_F0	.45	2	.22	.53	.596

- - - - -



(1-B-1) ANOVA results in Experiment 1-B:  
Speaker and Speech Rate as within-subjects factors

GET /FILE 'eratel.dat'.  
The SPSS/PC+ system file is read from  
file eratel.dat  
The file was created on 8/03/95 at 12:51:22  
and is titled SPSS/PC+ System File Written by Data Entry II  
The SPSS/PC+ system file contains  
20 cases, each consisting of  
99 variables (including system variables).  
99 variables will be used in this session.  
-----

MANOVA p1 p2 p3 p4 p5 p6 p7 p8 p9 p10  
/wsfactors speaker(2) rate(5) /wsdesign.  
  
20 cases accepted.  
0 cases rejected because of out-of-range factor values.  
0 cases rejected because of missing data.  
1 non-empty cells.  
  
0 design will be processed.

-----  
Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares						
Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	56.28	19	2.96			
CONSTANT	3179.29	1	3179.29	1073.32	.000	

-----  
Tests involving 'SPEAKER' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares						
Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	92.60	19	4.87			
SPEAKER	115.63	1	115.63	23.73	.000	

-----  
Tests involving 'RATE' Within-Subject Effect.

Mauchly sphericity test, W = .05059  
Chi-square approx. = 51.96966 with 9 D. F.  
Significance = .000  
  
Greenhouse-Geisser Epsilon = .47303  
Huynh-Feldt Epsilon = .52376  
Lower-bound Epsilon = .25000

AVERAGED Tests of Significance that follow multivariate tests are  
equivalent to  
univariate or split-plot or mixed-model approach to repeated measures.

Epsilons may be used to adjust d.f. for the AVERAGED results.

# EFFECT .. RATE

Multivariate Tests of Significance (S = 1, M = 1 , N = 7 )

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.74124	11.45831	4.00	16.00	.000
Hotellings	2.86458	11.45831	4.00	16.00	.000
Wilks	.25876	11.45831	4.00	16.00	.000
Roys	.74124				

- - - - -

Tests involving 'RATE' Within-Subject Effect.

AVERAGED Tests of Significance for P using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	101.24	76	1.33		
RATE	63.88	4	15.97	11.99	.000

- - - - -

Tests involving 'SPEAKER BY RATE' Within-Subject Effect.

Mauchly sphericity test, W = .26181  
 Chi-square approx. = 23.34066 with 9 D. F.  
 Significance = .005

Greenhouse-Geisser Epsilon = .57300  
 Huynh-Feldt Epsilon = .65598  
 Lower-bound Epsilon = .25000

AVERAGED Tests of Significance that follow multivariate tests are equivalent to univariate or split-plot or mixed-model approach to repeated measures. Epsilons may be used to adjust d.f. for the AVERAGED results.

# EFFECT .. SPEAKER BY RATE

Multivariate Tests of Significance (S = 1, M = 1 , N = 7 )

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.62100	6.55402	4.00	16.00	.003
Hotellings	1.63850	6.55402	4.00	16.00	.003
Wilks	.37900	6.55402	4.00	16.00	.003
Roys	.62100				

- - - - -

Tests involving 'SPEAKER BY RATE' Within-Subject Effect.

AVERAGED Tests of Significance for P using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	23.26	76	.31		
SPEAKER BY RATE	9.64	4	2.41	7.88	.000

- - - - -



**(1-B-2) ANOVA results in Experiment 1-B:**  
**Speaker and Speech Rate as within-subjects factors, and**  
**Sex of Listener (SEX) and Speech Rate Category of Listener (SR2CAT: slow,**  
**middle and fast) as between-subjects factors**

MANOVA p1 p2 p3 p4 p5 p6 p7 p8 p9 p10  
by sex(1,2) sr2cat(1,3) /wsfactors speaker(2) rate(5) /wsdesign.

20 cases accepted.  
0 cases rejected because of out-of-range factor values.  
0 cases rejected because of missing data.  
6 non-empty cells.

0 design will be processed.

- - - - -

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares						
Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	30.05	14	2.15			
CONSTANT	2408.12	1	2408.12	1121.81	.000	
SEX	1.65	1	1.65	.77	.396	
SR2CAT	24.27	2	12.14	5.65	.016	
SEX BY SR2CAT	11.29	2	5.65	2.63	.107	

- - - - -

Tests involving 'SPEAKER' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares						
Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	64.63	14	4.62			
SPEAKER	87.11	1	87.11	18.87	.001	
SEX BY SPEAKER	13.96	1	13.96	3.02	.104	
SR2CAT BY SPEAKER	1.20	2	.60	.13	.879	
SEX BY SR2CAT BY SPEAKER	2.48	2	1.24	.27	.768	

- - - - -

Tests involving 'RATE' Within-Subject Effect.

Mauchly sphericity test, W = .05412  
Chi-square approx. = 36.21364 with 9 D. F.  
Significance = .000

Greenhouse-Geisser Epsilon = .53184  
Huynh-Feldt Epsilon = .85380  
Lower-bound Epsilon = .25000

AVERAGED Tests of Significance that follow multivariate tests are equivalent to univariate or split-plot or mixed-model approach to repeated measures. Epsilons may be used to adjust d.f. for the AVERAGED results.

-----  
EFFECT .. SEX BY SR2CAT BY RATE  
Multivariate Tests of Significance (S = 2, M = 1/2, N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	1.03587	3.22325	8.00	24.00	.012
Hotellings	3.25454	4.06818	8.00	20.00	.005
Wilks	.18348	3.66997	8.00	22.00	.007
Roys	.73907				

- - - - -  
EFFECT .. SR2CAT BY RATE  
Multivariate Tests of Significance (S = 2, M = 1/2, N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.97865	2.87455	8.00	24.00	.021
Hotellings	3.40499	4.25623	8.00	20.00	.004
Wilks	.18897	3.57618	8.00	22.00	.008
Roys	.75733				

- - - - -  
EFFECT .. SEX BY RATE  
Multivariate Tests of Significance (S = 1, M = 1 , N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.43592	2.12518	4.00	11.00	.146
Hotellings	.77279	2.12518	4.00	11.00	.146
Wilks	.56408	2.12518	4.00	11.00	.146
Roys	.43592				

- - - - -  
EFFECT .. RATE  
Multivariate Tests of Significance (S = 1, M = 1 , N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.74780	8.15396	4.00	11.00	.003
Hotellings	2.96508	8.15396	4.00	11.00	.003
Wilks	.25220	8.15396	4.00	11.00	.003
Roys	.74780				

- - - - -  
Tests involving 'RATE' Within-Subject Effect.

AVERAGED Tests of Significance for P using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	51.97	56	.93		
RATE	41.79	4	10.45	11.26	.000
SEX BY RATE	4.03	4	1.01	1.08	.373
SR2CAT BY RATE	27.78	8	3.47	3.74	.001
SEX BY SR2CAT BY RAT	27.64	8	3.46	3.72	.001
E					

- - - - -  
Tests involving 'SPEAKER BY RATE' Within-Subject Effect.



Mauchly sphericity test, W = .10204  
Chi-square approx. = 28.33995 with 9 D. F.  
Significance = .001

Greenhouse-Geisser Epsilon = .51182  
Huynh-Feldt Epsilon = .81457  
Lower-bound Epsilon = .25000

AVERAGED Tests of Significance that follow multivariate tests are equivalent to univariate or split-plot or mixed-model approach to repeated measures. Epsilons may be used to adjust d.f. for the AVERAGED results.

EFFECT .. SEX BY SR2CAT BY SPEAKER BY RATE  
Multivariate Tests of Significance (S = 2, M = 1/2, N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.83449	2.14795	8.00	24.00	.071
Hotellings	2.71086	3.38858	8.00	20.00	.013
Wilks	.24741	2.77872	8.00	22.00	.027
Roys	.72088				

EFFECT .. SR2CAT BY SPEAKER BY RATE  
Multivariate Tests of Significance (S = 2, M = 1/2, N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.58829	1.25018	8.00	24.00	.314
Hotellings	1.25388	1.56734	8.00	20.00	.197
Wilks	.43385	1.42504	8.00	22.00	.241
Roys	.54787				

EFFECT .. SEX BY SPEAKER BY RATE  
Multivariate Tests of Significance (S = 1, M = 1 , N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.43834	2.14616	4.00	11.00	.143
Hotellings	.78042	2.14616	4.00	11.00	.143
Wilks	.56166	2.14616	4.00	11.00	.143
Roys	.43834				

EFFECT .. SPEAKER BY RATE  
Multivariate Tests of Significance (S = 1, M = 1 , N = 4 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.85421	16.11225	4.00	11.00	.000
Hotellings	5.85900	16.11225	4.00	11.00	.000
Wilks	.14579	16.11225	4.00	11.00	.000
Roys	.85421				

-----

Tests involving 'SPEAKER BY RATE' Within-Subject Effect.

AVERAGED Tests of Significance for P using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	18.69	56	.33		
SPEAKER BY RATE	10.78	4	2.69	8.07	.000
SEX BY SPEAKER BY RATE	1.31	4	.33	.98	.424
SR2CAT BY SPEAKER BY RATE	.92	8	.12	.35	.944
SEX BY SR2CAT BY SPEAKER BY RATE	2.84	8	.35	1.06	.402

- - - - -



(2-POLITENESS) ANOVA results in Experiment 2:  
First Style (Style\_1), Final Style (Style\_F) and Final Prosody (Pro\_F)  
as within-subjects factors

```
SPSS/PC+ The Statistical Package for IBM PC
6/10/96
SET WORKDEV=c.
SET WKSPACE=1500.

GET /FILE 'ff0nakp.dat'.
The SPSS/PC+ system file is read from
  file ff0nakp.dat
The file was created on  7/22/94 at 13:01:36
and is titled SPSS/PC+ System File Written by Data Entry II
The SPSS/PC+ system file contains
  19 cases, each consisting of
  37 variables (including system variables).
  37 variables will be used in this session.
-----

MANOVA p1 p2 p3 p4 p5 p6 p7 p8
  /wsfactors style_1(2) style_f(2) pro_f(2) /wsdesign.

    19 cases accepted.
    0 cases rejected because of out-of-range factor values.
    0 cases rejected because of missing data.
    1 non-empty cells.

    0 design will be processed.

- - - - -

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN CELLS              65.56      18      3.64
CONSTANT                 2769.69       1    2769.69    760.46      .000

- - - - -

Tests involving 'STYLE_1' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN CELLS              18.58      18      1.03
STYLE_1                   41.43       1    41.43    40.13      .000

- - - - -

Tests involving 'STYLE_F' Within-Subject Effect.

Tests of Significance for T3 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN CELLS              51.13      18      2.84
STYLE_F                  127.92       1    127.92    45.03      .000
```

- - - - -

Tests involving 'PRO\_F' Within-Subject Effect.

Tests of Significance for T4 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	15.03	18	.84		
PRO_F	4.45	1	4.45	5.32	.033

- - - - -

Tests involving 'STYLE\_1 BY STYLE\_F' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.83	18	.32		
STYLE_1 BY STYLE_F	7.77	1	7.77	23.97	.000

- - - - -

Tests involving 'STYLE\_1 BY PRO\_F' Within-Subject Effect.

Tests of Significance for T6 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.94	18	.33		
STYLE_1 BY PRO_F	1.94	1	1.94	5.87	.026

- - - - -

Tests involving 'STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T7 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.96	18	.33		
STYLE_F BY PRO_F	.31	1	.31	.93	.348

- - - - -

Tests involving 'STYLE\_1 BY STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T8 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.41	18	.30		
STYLE_1 BY STYLE_F B Y PRO_F	1.50	1	1.50	5.00	.038

- - - - -

**(2-ANGER) ANOVA results in Experiment 2:**  
**First Style (Style\_1), Final Style (Style\_F) and Final Prosody (Pro\_F)**  
**as within-subjects factors**

MANOVA a1 a2 a3 a4 a5 a6 a7 a8  
/wsfactors style\_1(2) style\_f(2) pro\_f(2) /wsdesign.



19 cases accepted.  
0 cases rejected because of out-of-range factor values.  
0 cases rejected because of missing data.  
1 non-empty cells.  
  
0 design will be processed.

- - - - -

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	65.15	18	3.62		
CONSTANT	2245.02	1	2245.02	620.30	.000

- - - - -

Tests involving 'STYLE\_1' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	31.18	18	1.73		
STYLE_1	35.56	1	35.56	20.53	.000

- - - - -

Tests involving 'STYLE\_F' Within-Subject Effect.

Tests of Significance for T3 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	58.80	18	3.27		
STYLE_F	277.67	1	277.67	85.00	.000

- - - - -

Tests involving 'PRO\_F' Within-Subject Effect.

Tests of Significance for T4 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	11.43	18	.64		
PRO_F	3.98	1	3.98	6.27	.022

- - - - -

Tests involving 'STYLE\_1 BY STYLE\_F' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	13.23	18	.74		
STYLE_1 BY STYLE_F	13.16	1	13.16	17.90	.001

- - - - -

Tests involving 'STYLE\_1 BY PRO\_F' Within-Subject Effect.

Tests of Significance for T6 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F

WITHIN CELLS	11.64	18	.65		
STYLE_1 BY PRO_F	2.95	1	2.95	4.55	.047

-----  
Tests involving 'STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T7 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	9.94	18	.55		
STYLE_F BY PRO_F	.22	1	.22	.40	.535

-----  
Tests involving 'STYLE\_1 BY STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T8 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	8.77	18	.49		
STYLE_1 BY STYLE_F B Y PRO_F	1.06	1	1.06	2.17	.158

-----

**(2-KINDNESS) ANOVA results in Experiment 2:**  
**First Style (Style\_1), Final Style (Style\_F) and Final Prosody (Pro\_F)**  
**as within-subjects factors**

MANOVA k1 k2 k3 k4 k5 k6 k7 k8  
/wsfactors style\_1(2) style\_f(2) pro\_f(2) /wsdesign.

19 cases accepted.  
0 cases rejected because of out-of-range factor values.  
0 cases rejected because of missing data.  
1 non-empty cells.  
  
0 design will be processed.

-----

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	61.99	18	3.44		
CONSTANT	1757.94	1	1757.94	510.43	.000

-----

Tests involving 'STYLE\_1' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	23.22	18	1.29		



STYLE_1	27.39	1	27.39	21.23	.000
---------	-------	---	-------	-------	------

-----  
Tests involving 'STYLE\_F' Within-Subject Effect.

Tests of Significance for T3 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	32.86	18	1.83		
STYLE_F	178.94	1	178.94	98.03	.000

-----  
Tests involving 'PRO\_F' Within-Subject Effect.

Tests of Significance for T4 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	9.53	18	.53		
PRO_F	10.28	1	10.28	19.40	.000

-----  
Tests involving 'STYLE\_1 BY STYLE\_F' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	12.10	18	.67		
STYLE_1 BY STYLE_F	26.91	1	26.91	40.03	.000

-----  
Tests involving 'STYLE\_1 BY PRO\_F' Within-Subject Effect.

Tests of Significance for T6 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	9.22	18	.51		
STYLE_1 BY PRO_F	1.60	1	1.60	3.12	.094

-----  
Tests involving 'STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T7 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	12.12	18	.67		
STYLE_F BY PRO_F	1.00	1	1.00	1.48	.239

-----  
Tests involving 'STYLE\_1 BY STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T8 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5.37	18	.30		
STYLE_1 BY STYLE_F B Y PRO_F	2.31	1	2.31	7.73	.012

-----

(2-NATURALNESS) ANOVA results in Experiment 2:  
First Style (Style\_1), Final Style (Style\_F) and Final Prosody (Pro\_F)  
as within-subjects factors

```
MANOVA n1 n2 n3 n4 n5 n6 n7 n8
/wsufactors style_1(2) style_f(2) pro_f(2) /wsdesign.

19 cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
1 non-empty cells.

0 design will be processed.

- - - - -

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN CELLS              185.47      18      10.30
CONSTANT                  3169.57       1     3169.57     307.61      .000

- - - - -

Tests involving 'STYLE_1' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN CELLS              70.32      18       3.91
STYLE_1                   64.92       1     64.92     16.62      .001

- - - - -

Tests involving 'STYLE_F' Within-Subject Effect.

Tests of Significance for T3 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN CELLS              35.20      18       1.96
STYLE_F                   23.26       1     23.26     11.90      .003

- - - - -

Tests involving 'PRO_F' Within-Subject Effect.

Tests of Significance for T4 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN CELLS              16.42      18       .91
PRO_F                     2.37       1      2.37      2.60      .124

- - - - -

Tests involving 'STYLE_1 BY STYLE_F' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F
```



WITHIN CELLS	31.03	18	1.72		
STYLE_1 BY STYLE_F	5.68	1	5.68	3.29	.086

-----

Tests involving 'STYLE\_1 BY PRO\_F' Within-Subject Effect.

Tests of Significance for T6 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
---------------------	----	----	----	---	----------

WITHIN CELLS	11.47	18	.64		
STYLE_1 BY PRO_F	4.16	1	4.16	6.53	.020

-----

Tests involving 'STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T7 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
---------------------	----	----	----	---	----------

WITHIN CELLS	5.84	18	.32		
STYLE_F BY PRO_F	.08	1	.08	.24	.632

-----

Tests involving 'STYLE\_1 BY STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T8 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
---------------------	----	----	----	---	----------

WITHIN CELLS	7.73	18	.43		
STYLE_1 BY STYLE_F B Y PRO_F	.12	1	.12	.28	.601

-----

(2-REACTION TIME) ANOVA results in Experiment 2:  
First Style (Style\_1), Final Style (Style\_F) and Final Prosody (Pro\_F)  
as within-subjects factors

GET /FILE 'ff0time.dat'.  
The SPSS/PC+ system file is read from  
file ff0time.dat  
The file was created on 7/22/94 at 13:27:26  
and is titled SPSS/PC+ System File Written by Data Entry II  
The SPSS/PC+ system file contains  
19 cases, each consisting of  
13 variables (including system variables).  
13 variables will be used in this session.

MANOVA t1 t2 t3 t4 t5 t6 t7 t8  
/wsfactors style\_1(2) style\_f(2) pro\_f(2) /wsdesign.

19 cases accepted.  
0 cases rejected because of out-of-range factor values.  
0 cases rejected because of missing data.  
1 non-empty cells.  
  
0 design will be processed.

-----  
Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	2142.46	18	119.03		
CONSTANT	55856.78	1	55856.78	469.28	.000

-----  
Tests involving 'STYLE\_1' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	141.62	18	7.87		
STYLE_1	45.76	1	45.76	5.82	.027

-----  
Tests involving 'STYLE\_F' Within-Subject Effect.

Tests of Significance for T3 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	99.26	18	5.51		
STYLE_F	67.38	1	67.38	12.22	.003

-----  
Tests involving 'PRO\_F' Within-Subject Effect.

Tests of Significance for T4 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
---------------------	----	----	----	---	----------



WITHIN CELLS	48.81	18	2.71		
PRO_F	.32	1	.32	.12	.734

- - - - -

Tests involving 'STYLE\_1 BY STYLE\_F' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	188.84	18	10.49		
STYLE_1 BY STYLE_F	47.76	1	47.76	4.55	.047

- - - - -

Tests involving 'STYLE\_1 BY PRO\_F' Within-Subject Effect.

Tests of Significance for T6 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	93.81	18	5.21		
STYLE_1 BY PRO_F	30.24	1	30.24	5.80	.027

- - - - -

Tests involving 'STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T7 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	100.86	18	5.60		
STYLE_F BY PRO_F	7.25	1	7.25	1.29	.270

- - - - -

Tests involving 'STYLE\_1 BY STYLE\_F BY PRO\_F' Within-Subject Effect.

Tests of Significance for T8 using UNIQUE sums of squares					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	65.60	18	3.64		
STYLE_1 BY STYLE_F B Y PRO_F	11.61	1	11.61	3.18	.091

- - - - -

# **APPENDIX A**

## **Scenarios given to the speakers in the recording sessions**

A.1. Scenarios 1 and 2 (in Japanese)

A.2. English translation of A.1



## A.1. Scenarios 1 and 2 (in Japanese)

## 場面1 空港の税関

登場人物: 税関の人 (Speaker) → 乗客 A, B, C

20代で、この仕事について1年くらいの人  
(i.e., 慣れきったベテランではない)

乗客 A: きちんとした身なりの初老の紳士  
外務省の高官であることを税関の人は知っているが  
個人的面識はない

B: カジュアルな身なりの若い学生

C: 少しおぼろげにしている、あまりまともでない身なりの中年男性  
他の乗客にからんだりして、列をみだしている

task (目的: 役になりきるために)

税関の人: 乗客が「どの国に/どのくらい/何のために  
行っていたかを質問する

→ 乗客 A: 敬意をもって/少しかしこまって

B: 気楽な調子で

C: 権威をもって(自分の方が立場が上という感じで)

乗客 A: アメリカに 1week, アメリカの政府高官との会議出席のため  
フランスに 1week, 国際学会出席のため

B: ハワイに 1week, スモークダズイングをするため

C: 韓国に 3日間, 社員旅行で  
(additional)

税関の人 → 乗客 C

他の乗客に迷惑をかけないように注意する。

Fixed line (マイクに向かって、税関の人になりきって次の文を数回)  
しゃべる

「荷物はこれだけですか」

## 場面2 電話での会話

市役所の「市民サービス課」にて

登場人物: 市役所の人 (Speaker) → 市民 A, B, C

20代で、この仕事について1年くらいの人

市民 A: この地方の名士で、感じのよい立派な初老の紳士  
大企業の役員を、去年退いた  
市の計画している市民病院増築に寄付を申し出ている

B: カジュアルな身なりの若い学生

C: 少しよぼろっているようにみえる、あまりまともでない身なり  
の中年男性。仮は前にも、何回かささいな事で文句を  
言いに来たり、変な質問をしたりして、一種のトラブル  
メーカーである。丁度いいに待っていると、いつまでも  
食い下かって仁事にならない。

。Task (目的: 役になりきるために)

市民 (A/B/C) が市民サービス課に来ている

dialogue A:

市民 A: 市民病院増築に、何千万円かを寄付することを  
考えている。市のプラン、寄付のための手続きなどを知りたい。

市役所の人: 今検討中で、建築は来年の4月の予定。  
詳しいことをお話しするために、来週、上司と伺いたい  
都合のよい日を教えてほしい。

市民 A: 手帳がないので、今返事ができない  
後日電話してほしい。

市役所の人: 後で電話する。  
わざわざ来てくれたことのお礼。



dialogue B:

市民 B: 市民ホールを1月23日に使いたいかあっているかを知りたい

市役所の人: コンビニで調べるので、ちょっと待ってほしい  
コンビニがダウンしているので、後で電話する

市民 B: よろしくお願いします。

dialogue C:

市民 C: となりのピアノの音がうるさい。何とかならないか。  
赤城坊も犬も夜にならなくなるので何日もよく眠れない

市役所の人: またですか。  
それはいつのことですか

市民 C: ここ 2~3ヶ月ずっと。特に先週はひどかった..  
....

(文句がだらだら続く)

市役所の人: わかりました。わかりました。  
隣人と話をした上、また電話する旨、告げる。

・ Fixed line (マイクに向かって市役所の人になりきって次の文を  
数回しゃべる)

電話で

「もしもし、赤城さんのお宅ですか」

To 市民 A: 敬意をもって (少しかしこまって)

B: 気楽な調子で

C: 権威をもって (自分の方が立場が上という感じで)

## A.2. English translation of A.1

=====

Scenario 1: at a customs counter at an airport

=====

Speaker: a customs officer talking to Passenger A/B/C

He is in his twenties, and has been in this job for a year (so he is not experienced)

Passenger A: a finely dressed gentleman

The customs officer knows that this gentleman is a high official of the Department of Foreign Affairs, but is not personally acquainted with him.

Passenger B: a casually dressed young student

Passenger C: a shabby looking middle-aged drunk, who has been picking a quarrel with other passengers

-----

Task (in order to get into the roles)

-----

The customs officer asks passengers which countries they went to, how long and for what purpose they were there: speak to Passenger A with respect/slightly formally; to Passenger B, casually or uninhibitedly; to Passenger C, with authority with a definite attitude.

Answers:

Passenger A: visited the USA for 1 week for a meeting with high American officials; visited France for 1 week for an international conference

Passenger B: visited Hawaii for 1 week for scuba diving

Passenger C: visited Korea for 3 days as part of his company recreation tour  
(additional task) The officer tells Passenger C not to trouble other passengers in the queue.



-----  
Fixed line (put yourself into the role of the customs officer, and speak the following line several times to the microphone)  
-----

Nimotsu-wa koredake desuka  
(‘Is this all the luggage you have?’)

=====

Scenario 2: at a citizen service counter at a local government office

=====

Speaker: a public officer talking to Citizen A/B/C  
He is in his twenties, and has been in this job for a year (so he is not experienced)

Citizen A: a finely dressed, nice gentleman  
He is a local celebrity. He retired from a position of an executive at a big company last year, and is offering a donation to the extension work for a city hospital.

Citizen B: a casually dressed young student

Citizen C: a shabby looking middle-aged drunk  
He is known as a trouble-maker, who came to the office several times before, making complaints about trifling matters and asking annoying questions. If you treat him nicely, he does not go away and will continue to disturb your work.

-----

Task (in order to get into the roles)

-----

Citizen A/B/C are at the counter.

/\*\*\*\*\*/

A sample dialogue with Citizen A:

/\*\*\*\*\*/

A: I am thinking of making a donation (tens of millions of yen) to the hospital extension project and would like to know the city plans and procedures, etc.

Public officer: The plan is now under final review. The construction is to begin next April. I would like to visit you about this matter with my boss next week, so I would like to know when would be convenient for you.

A: I cannot give you a definite date without my diary. Please call me later.

Public officer: We'll call you later.

He expresses his gratitude to the gentleman.

/\*\*\*\*\*/

A sample dialogue with Citizen B:

/\*\*\*\*\*/

B: I want to use the city hall on the 23rd of January. Is it available?

Public officer: I'll check it with a computer. Wait a minute, please. .... The system is now down. I'll call you later.

B: Thanks a lot.

/\*\*\*\*\*/

A sample dialogue with Citizen C:

/\*\*\*\*\*/

C: The piano practice of my neighbours is so noisy! Can you please do something to stop it? The baby keeps crying and the dog keeps barking especially at night, and I haven't slept for days!!

Public officer: What again? How long has it been going this time?

C: for the last couple of months. It was absolutely terrible last week!  
..... (continues complaining)

Public Officer: All right, all right. I'll talk to your neighbours and call you later.



-----

Fixed line (put yourself into the role of the public officer, and speak the following line several times to the microphone)

-----

Moshimoshi Akagi-san no otaku desuka.  
('Hello, is that Mr. Akagi speaking?')

# APPENDIX B

**Instructions and a part of the answer sheet  
for utterance evaluation**



## INSTRUCTIONS and ANSWER SHEET FOR UTTERANCE EVALUATION

SEX: Male Female

AGE: \_\_\_\_\_

HOMETOWN: \_\_\_\_\_

-----

Practice session

-----

[P1]            1st            2nd

[P2]            1st            2nd

-----

You are going to hear only two sentences in this session.

1. Nimotsu wa koredake desuka
2. Moshimoshi akagi-san no otaku desuka

The speakers were given some scenarios describing the situations, and asked to say these two sentences appropriately in a given situation.

1. Nimotsu wa koredake desuka ('Nimotsu' sentence)

Setting: at a customs counter at an airport

Speaker: a customs officer

Addressee:

- (A): a respectable gentleman
- (B): a young, casually-dressed student
- (C): a drunk/trouble-maker

2. Moshimoshi akagi-san no otaku desuka ('Moshi' sentence)

Setting: telephone conversation

Speaker: a public officer at a local government office

Addressee:

- (A): a gentleman who is thinking of making  
a substantial donation to the city's project
- (B): a young student who wants to book the city hall
- (C): a drunk/trouble-maker,  
who is always making complaints about something

There are 6 subsections on this tape:

-----  
A: 32 pairs x 2 sentences  
-----

Addressee: (A): Gentleman  
Tone: politely ('teineini; keii o motte')

A-1: 'Nimotsu' sentence  
A-2: 'Moshi' sentence

-----  
B: 32 pairs x 2 sentences  
-----

Addressee: (B): Student  
Tone: casually ('kirakuni')

B-1: 'Nimotsu' sentence  
B-2: 'Moshi' sentence

-----  
C: 32 pairs x 2 sentences  
-----

Addressee: (C): Drunk / Trouble-maker  
Tone: authoritative-casually  
('kirakuni, demo tsuyoi taidode')

C-1: 'Nimotsu' sentence  
C-2: 'Moshi' sentence

-----  
Each pair consists of

- beep
- Utterance 1
- Utterance 2
- 2-second silence

Each subsection consists of 32 pairs, and 2 beep tones at the end.

#### INSTRUCTION:

Encircle the utterance (1st or 2nd), which sounded more polite for Section A, more casual for Section B, and more authoritative-casual for Section C, to you.



-----  
Section A-1: Speak **POLITELY**  
-----

- |      |     |     |
|------|-----|-----|
| [ 1] | 1st | 2nd |
| [ 2] | 1st | 2nd |
| [ 3] | 1st | 2nd |
| [ 4] | 1st | 2nd |
| [ 5] | 1st | 2nd |
| [ 6] | 1st | 2nd |
| [ 7] | 1st | 2nd |
| [ 8] | 1st | 2nd |
| [ 9] | 1st | 2nd |
| [10] | 1st | 2nd |
| [11] | 1st | 2nd |
| [12] | 1st | 2nd |
| [13] | 1st | 2nd |
| [14] | 1st | 2nd |
| [15] | 1st | 2nd |
| [16] | 1st | 2nd |

(continued A-1)

[17]	1st	2nd
[18]	1st	2nd
[19]	1st	2nd
[20]	1st	2nd
[21]	1st	2nd
[22]	1st	2nd
[23]	1st	2nd
[24]	1st	2nd
[25]	1st	2nd
[26]	1st	2nd
[27]	1st	2nd
[28]	1st	2nd
[29]	1st	2nd
[30]	1st	2nd
[31]	1st	2nd
[32]	1st	2nd



# APPENDIX C

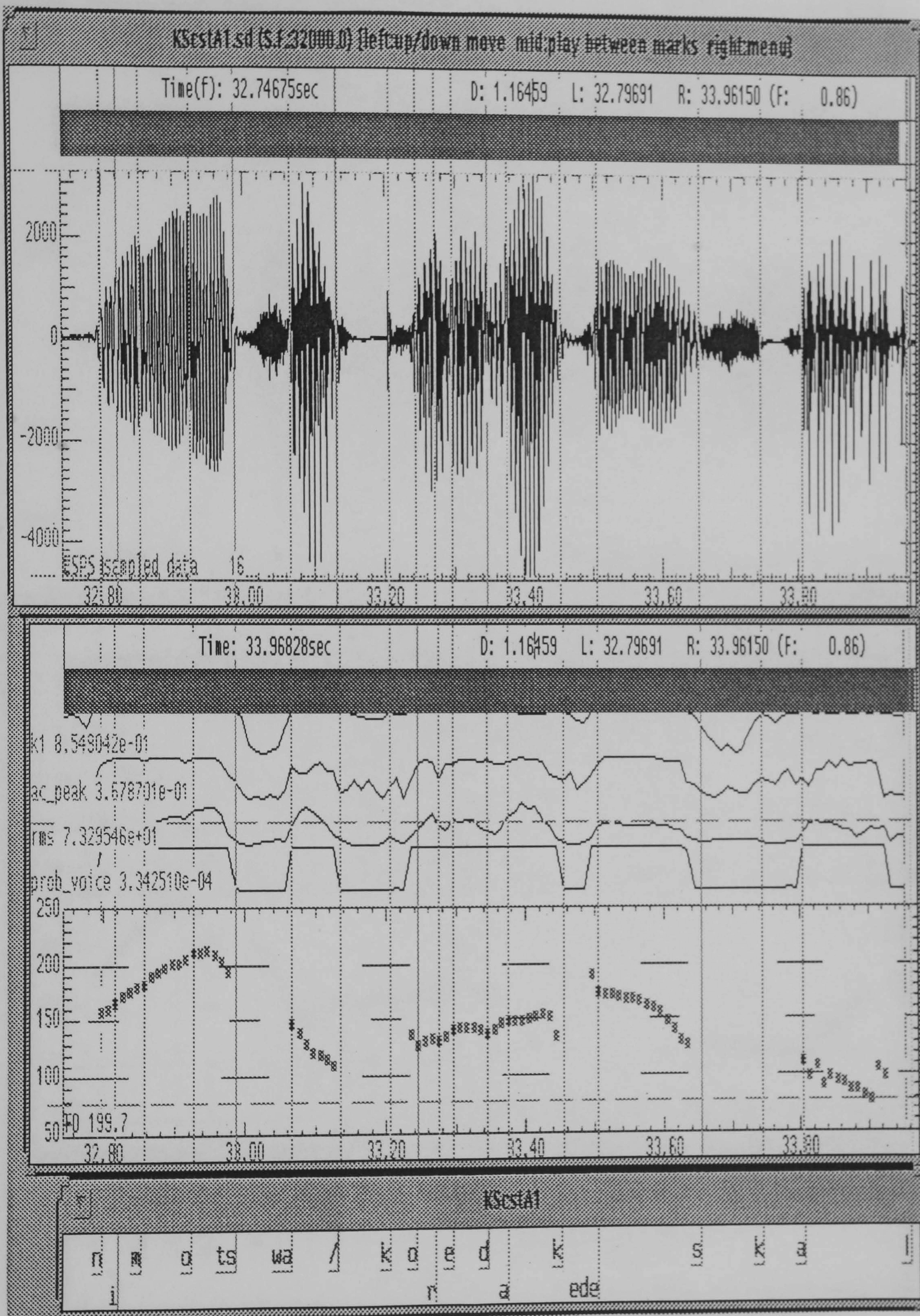
**Waveforms and f0 contours of two sentences spoken by three male speakers in a polite and casual manner**

C.1. 'Luggage' sentence

C.2. 'Hello' sentence

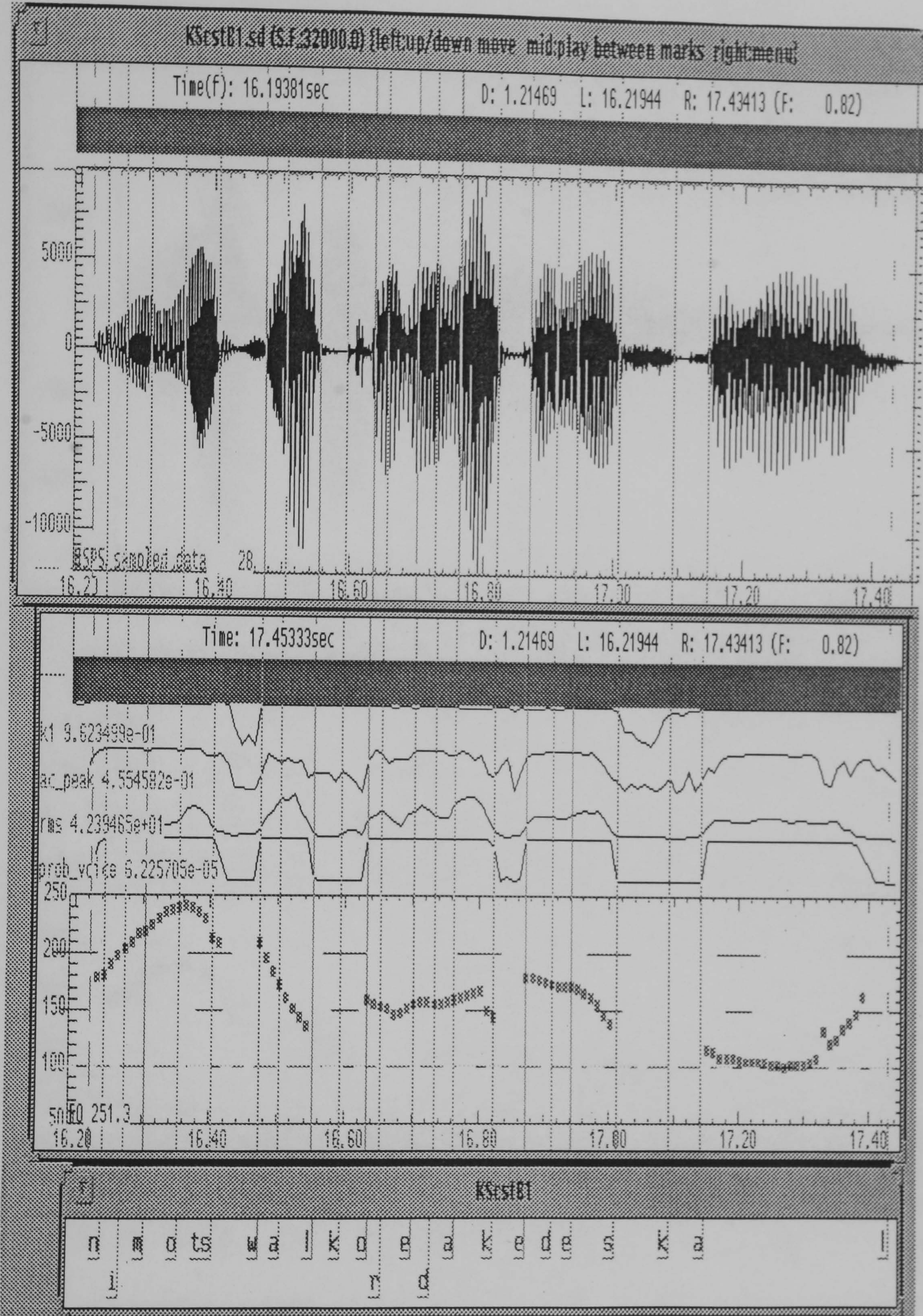


## C.1.KS-polite: KS's POLITE utterance of the 'LUGGAGE' sentence



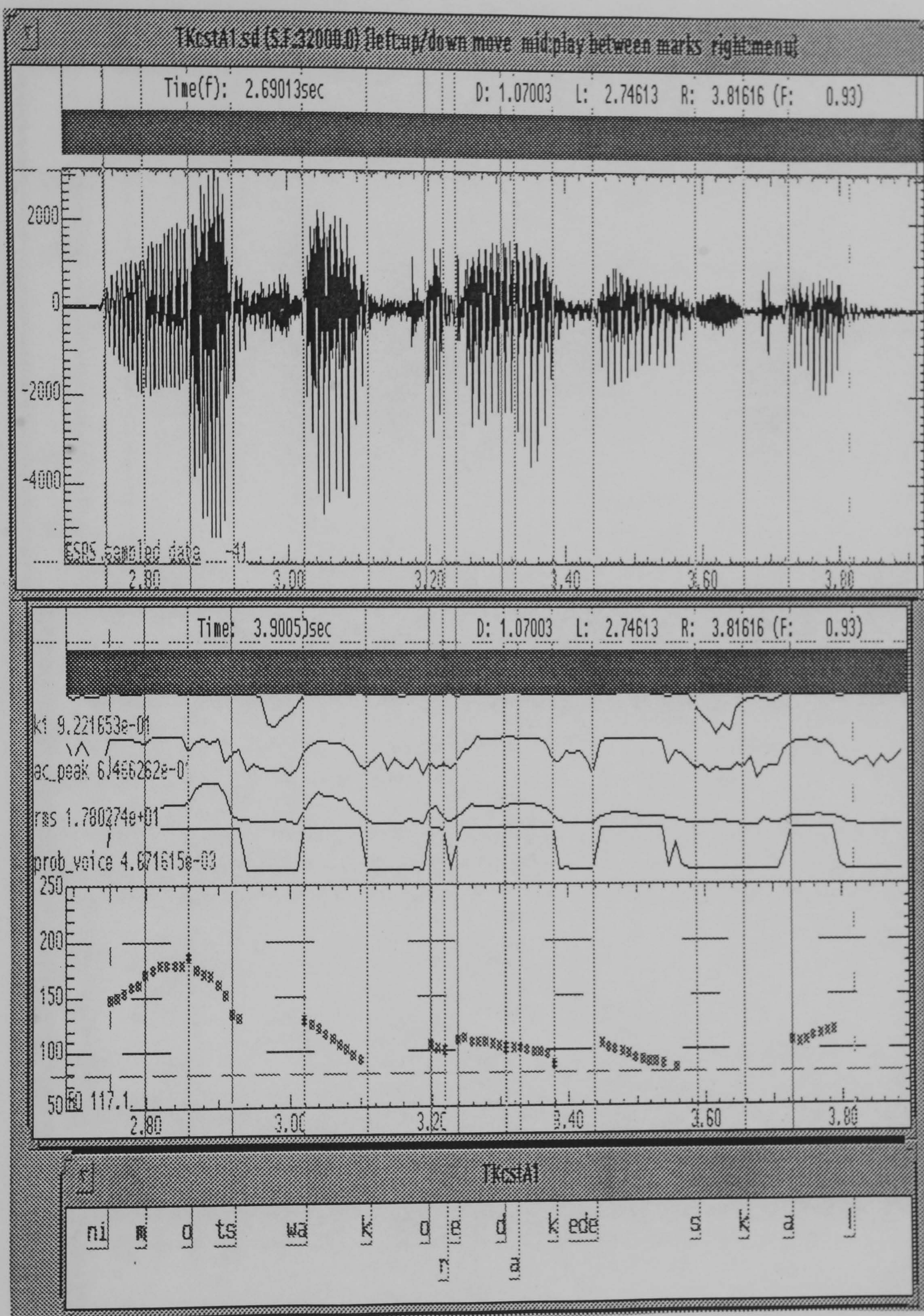


C.1.KS-casual: KS's CASUAL utterance of the 'LUGGAGE' sentence



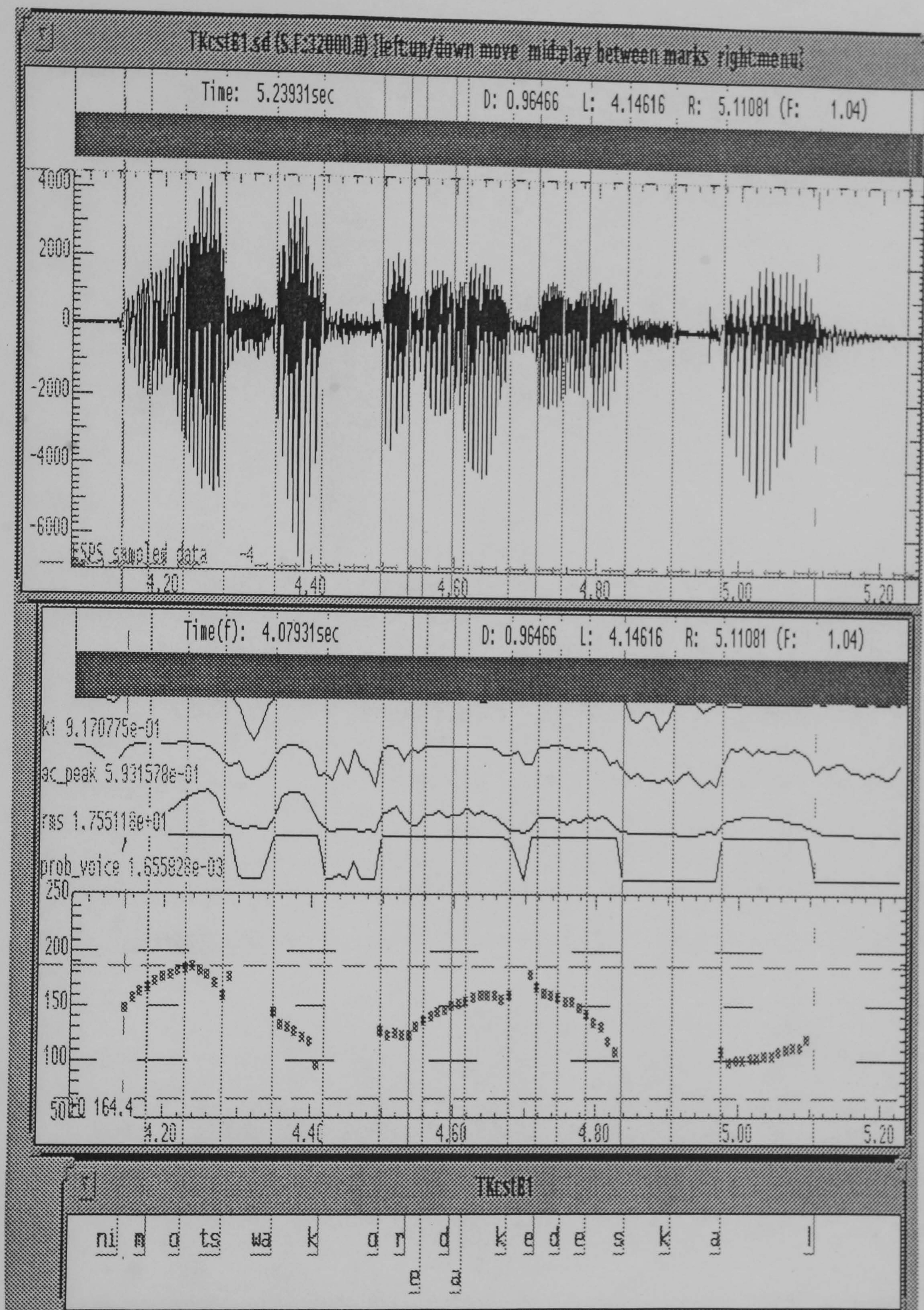


## C.1.TK-polite: TK's POLITE utterance of the 'LUGGAGE' sentence



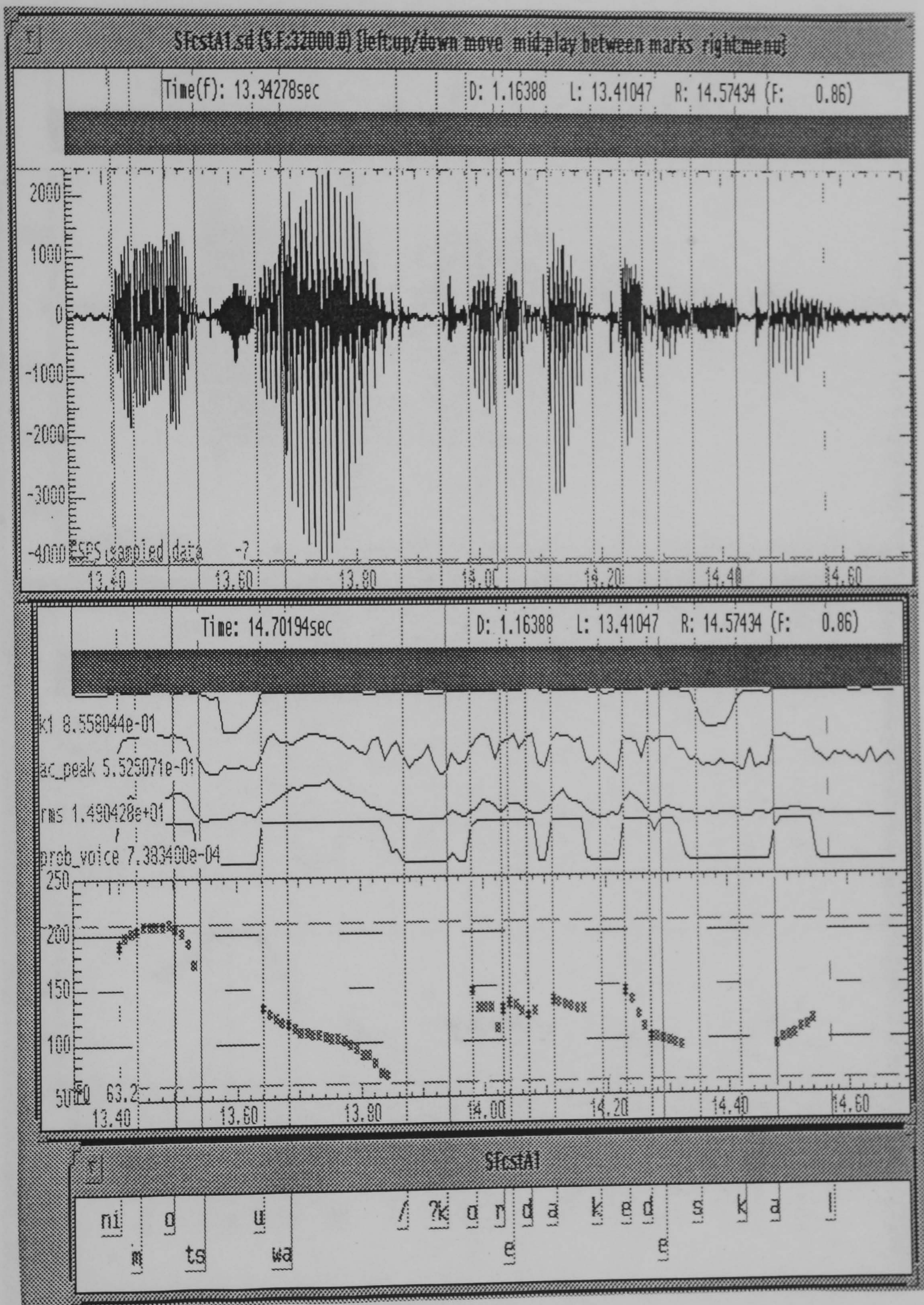


# C.1.TK-casual: TK's CASUAL utterance of the 'LUGGAGE' sentence



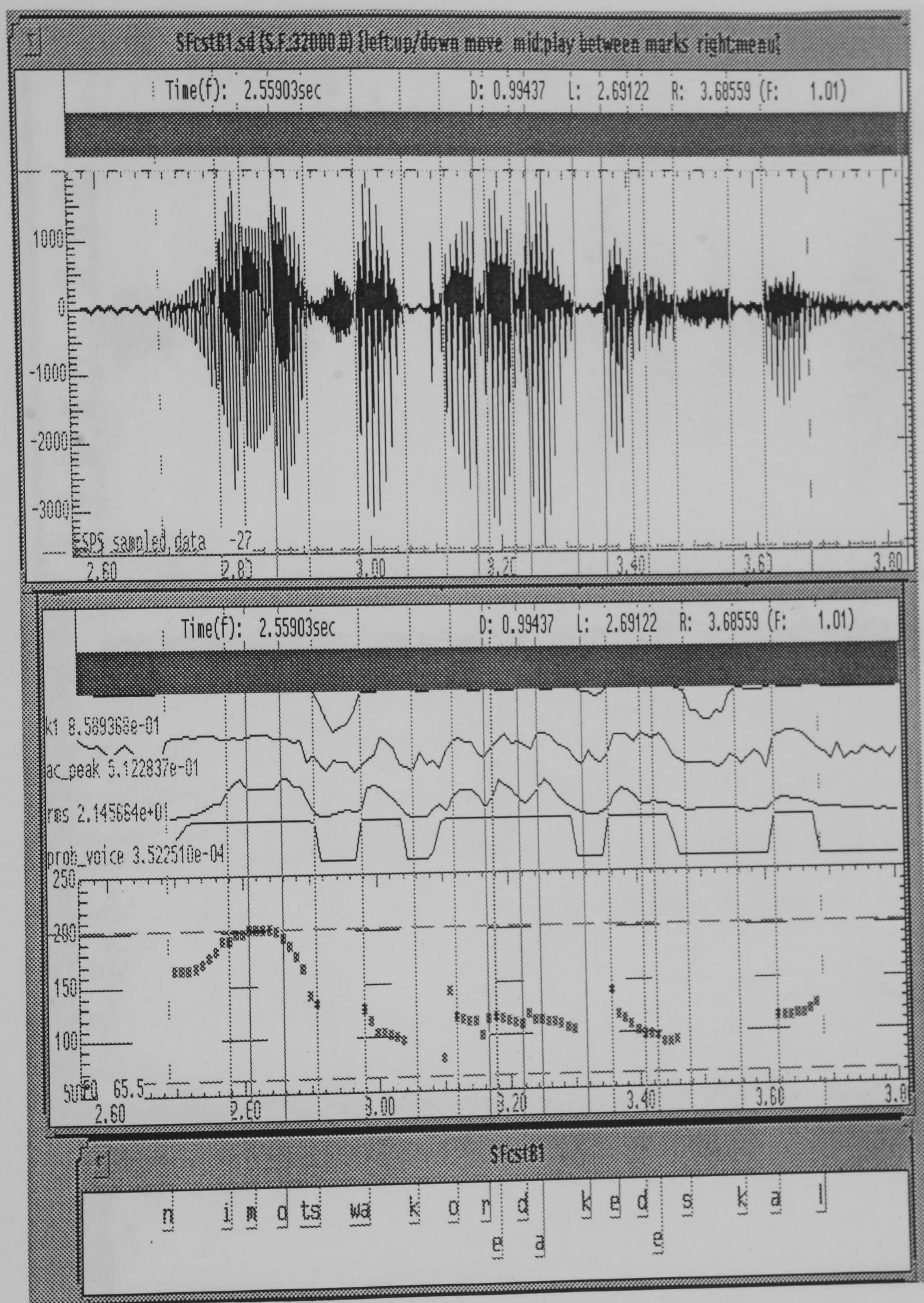


## C.1.SF-polite: SF's POLITE utterance of the 'LUGGAGE' sentence



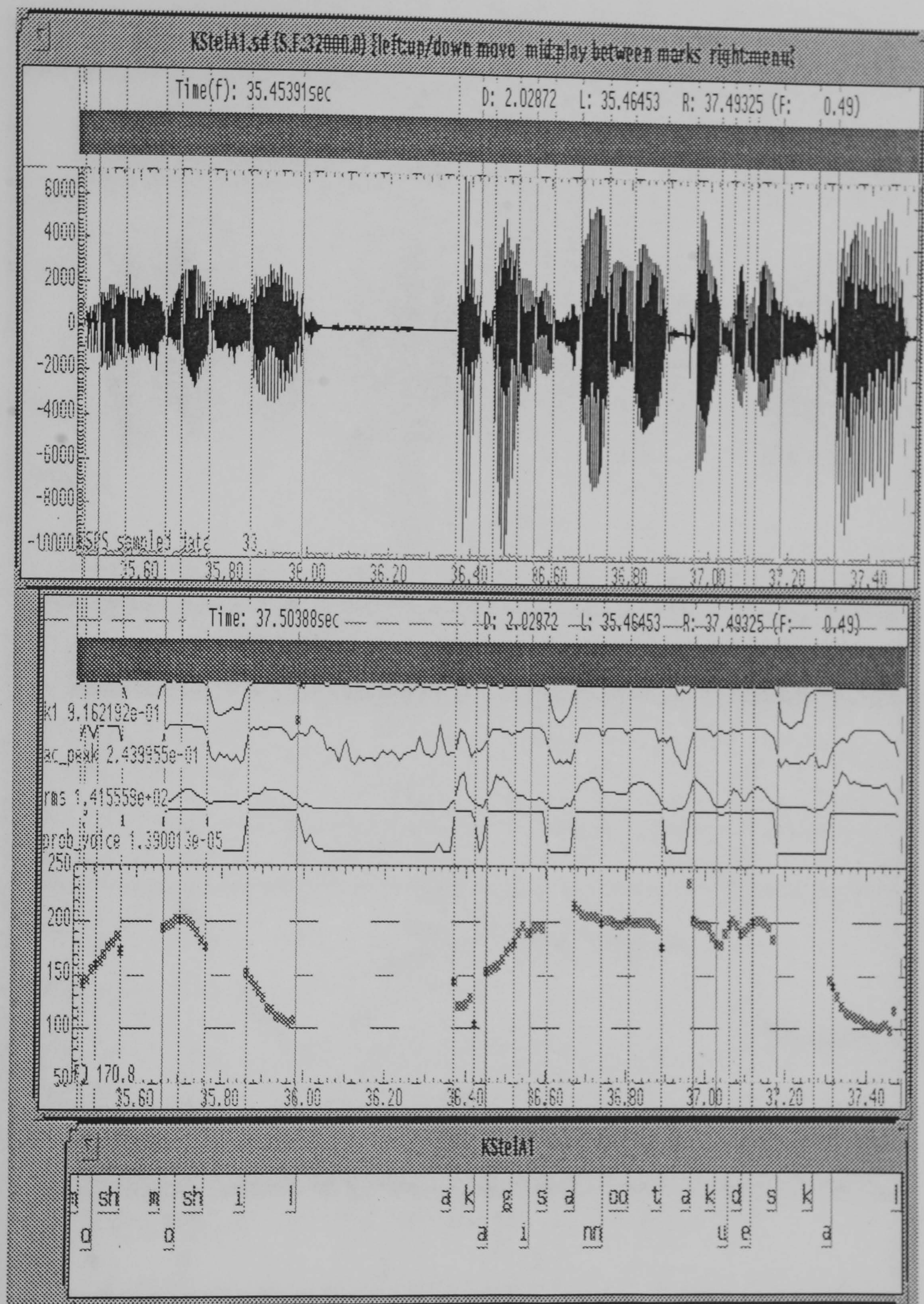


## C.1.SF-casual: SF's CASUAL utterance of the 'LUGGAGE' sentence



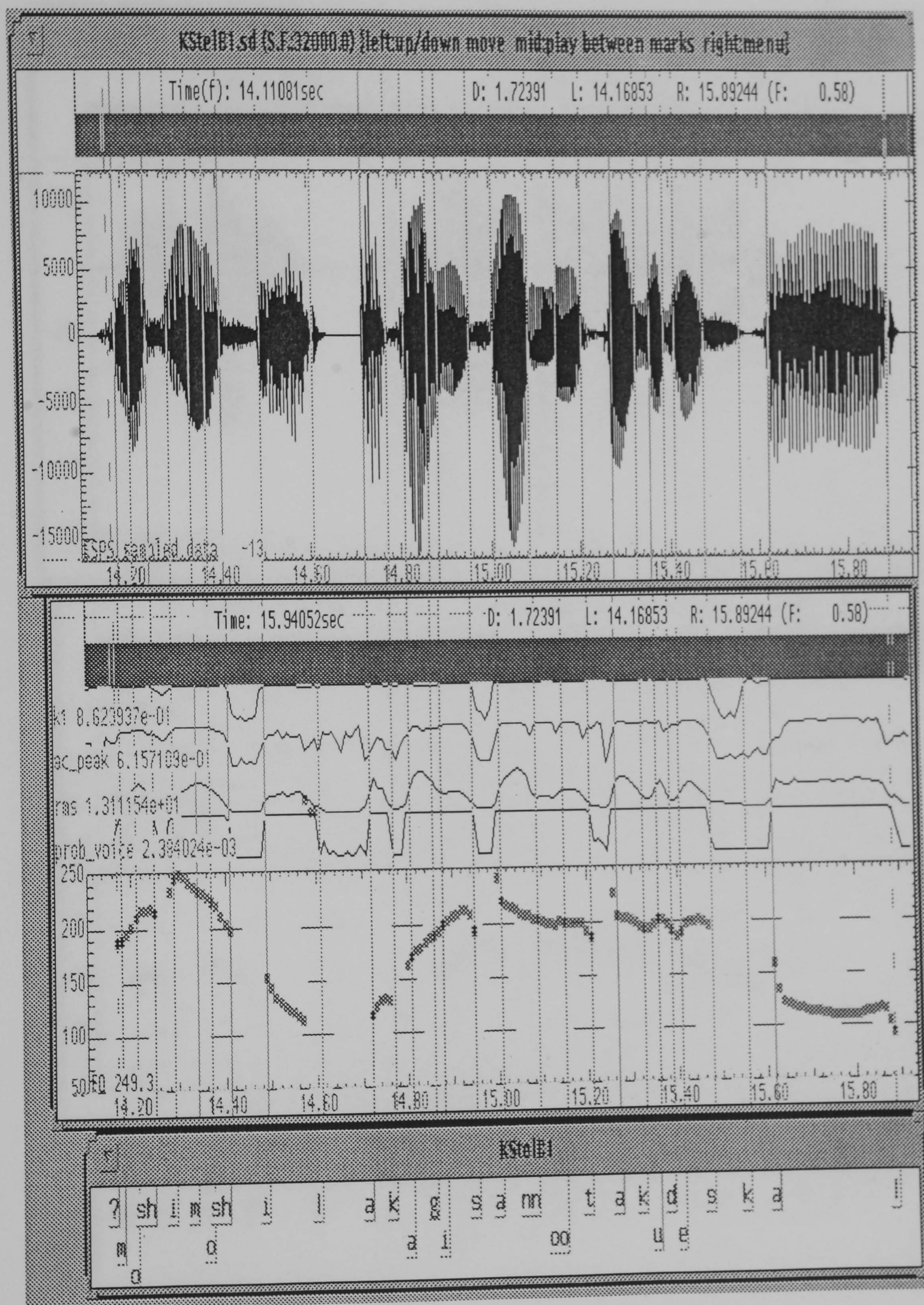


# C.2.KS-polite: KS's POLITE utterance of the 'HELLO' sentence



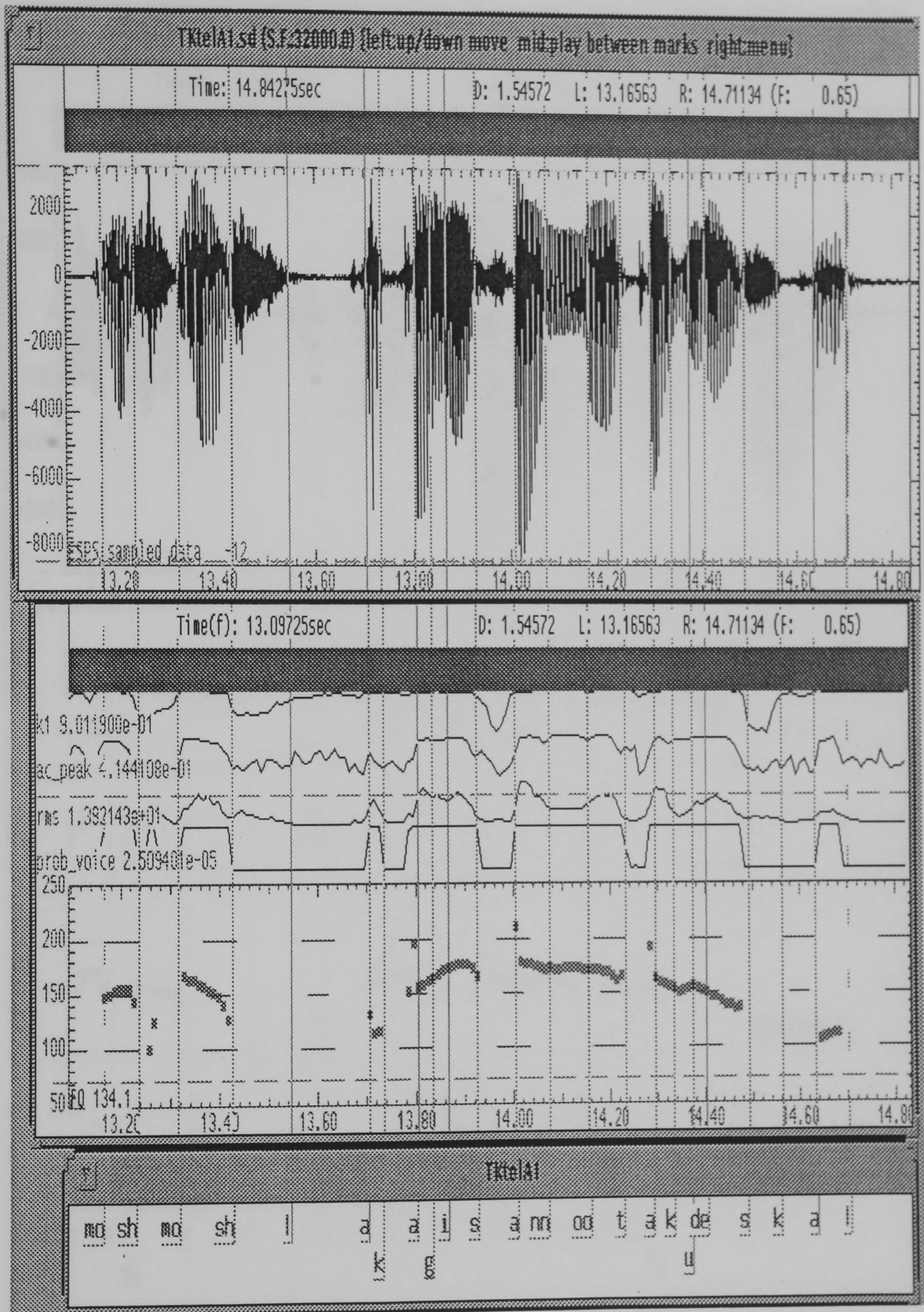


## C.2.KS-casual: KS's CASUAL utterance of the 'HELLO' sentence



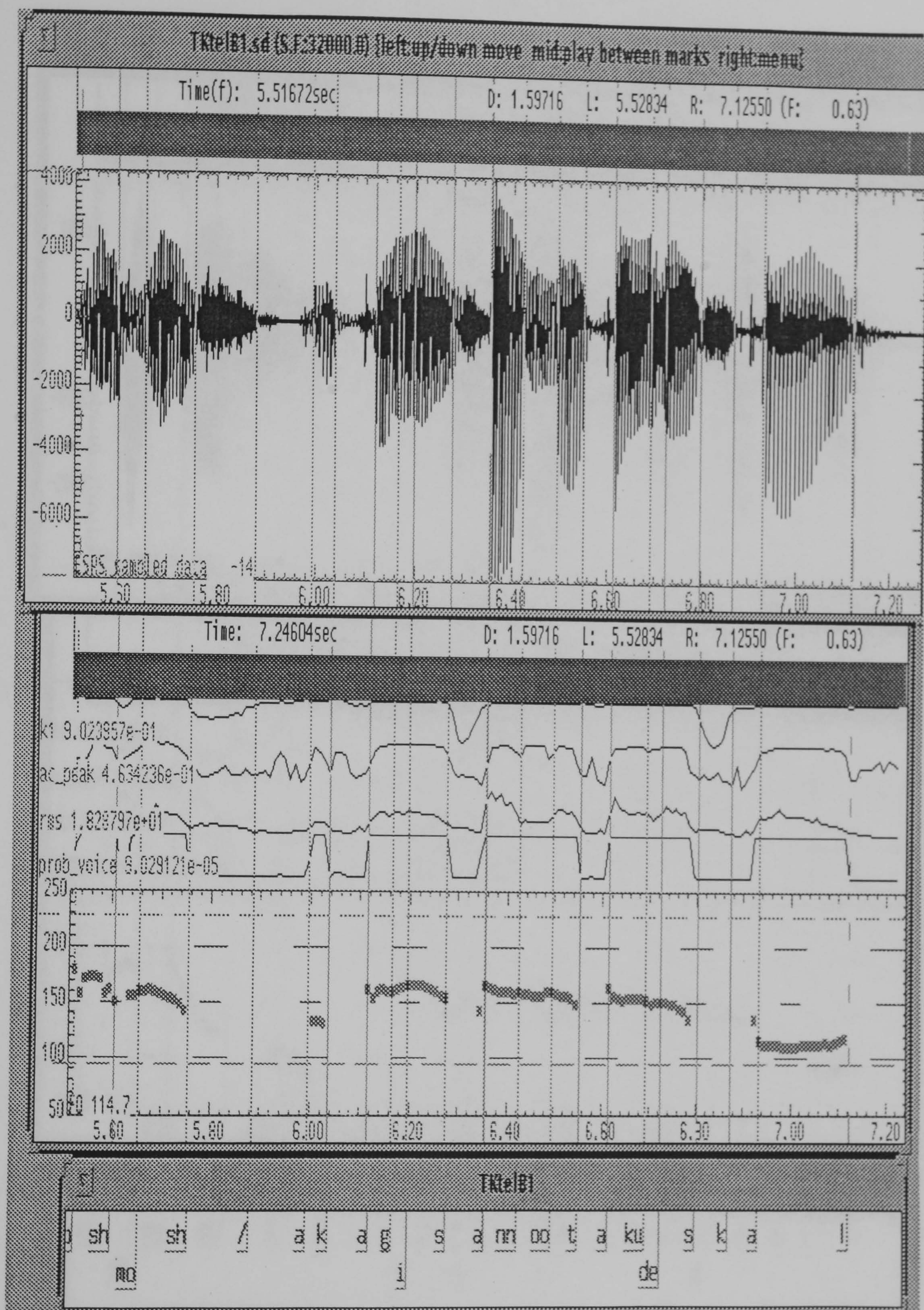


## C.2.TK-polite: TK's POLITE utterance of the 'HELLO' sentence



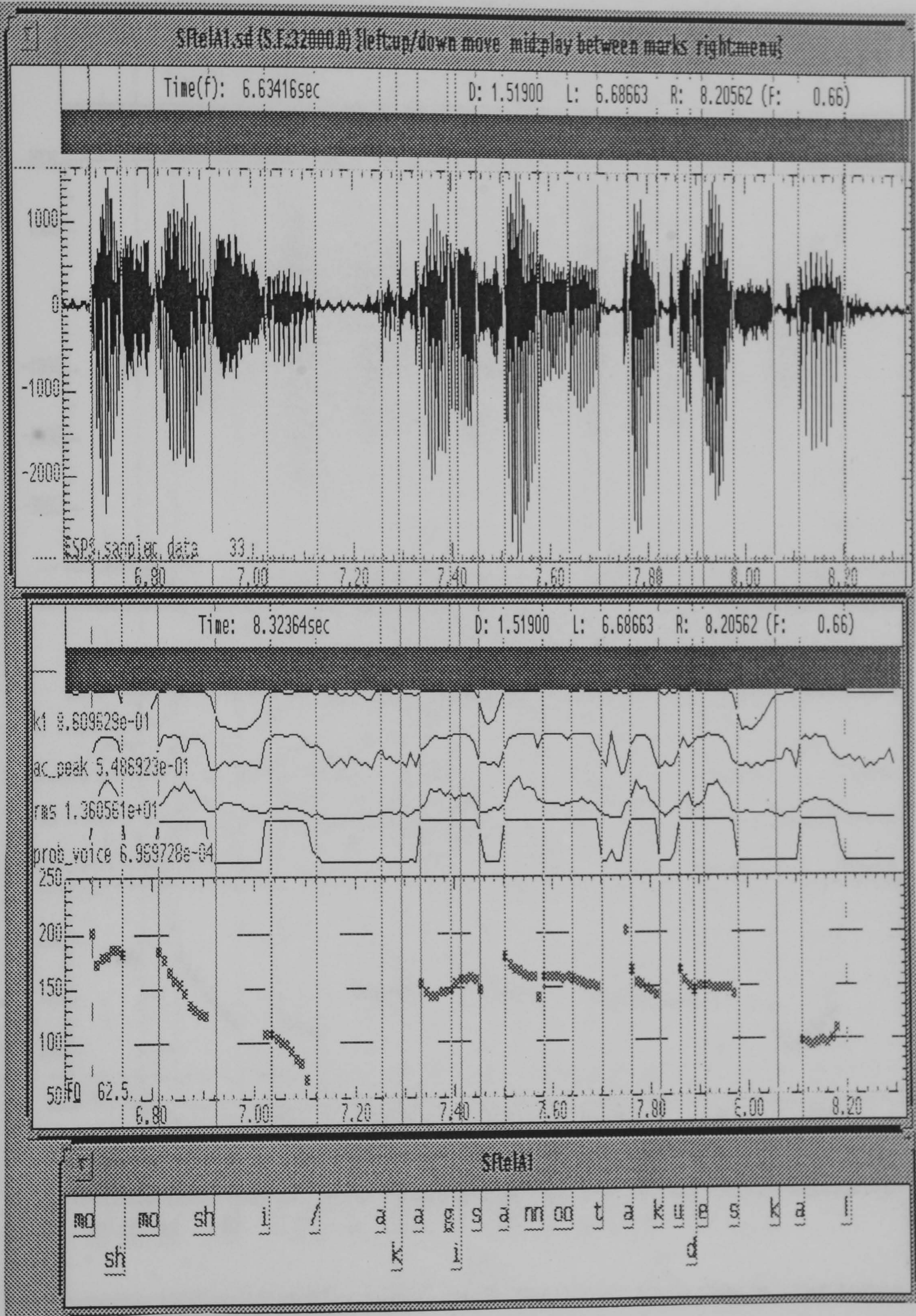


# C.2.TK-casual: TK's CASUAL utterance of the 'HELLO' sentence



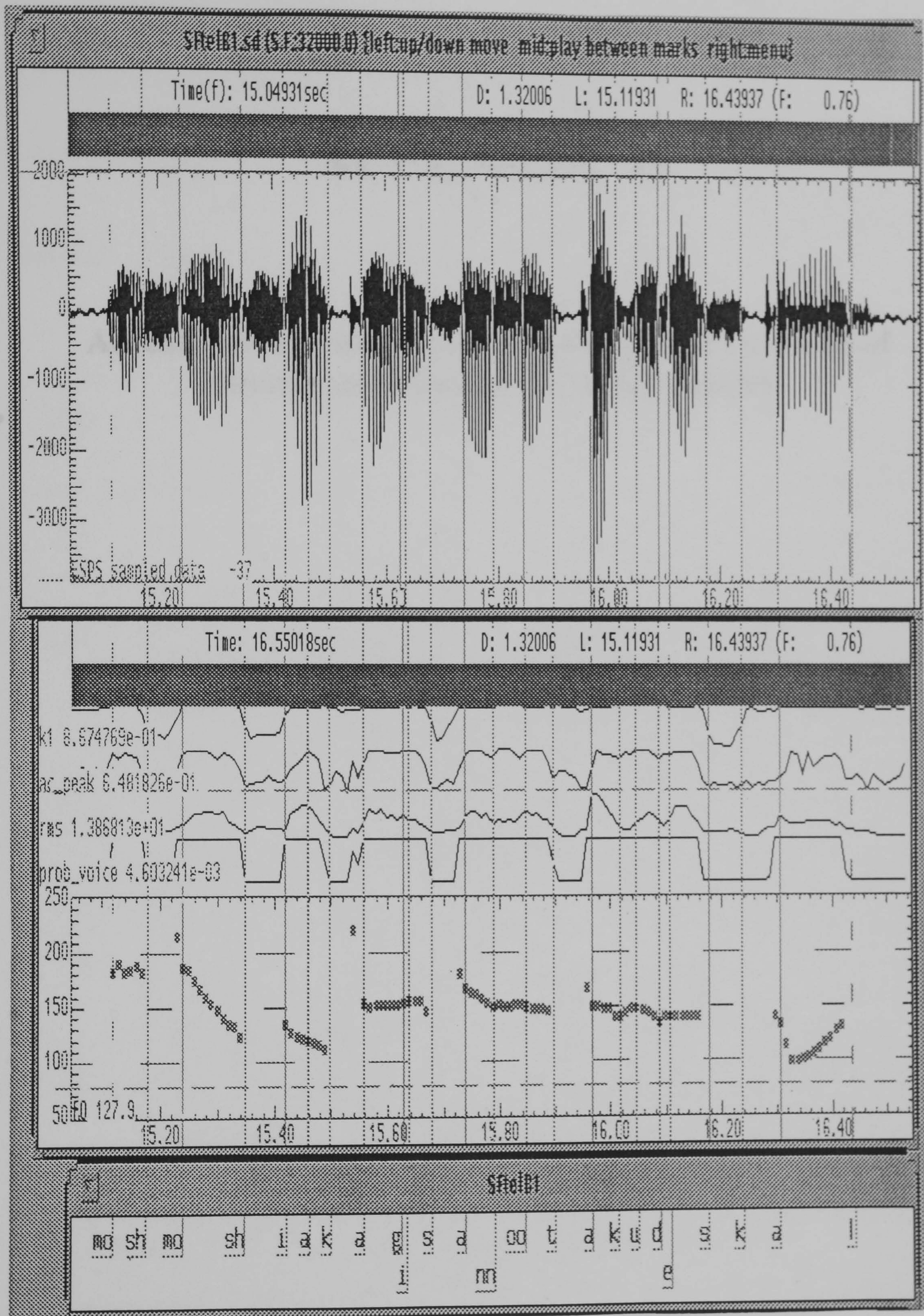


C.2.SF-polite: SF's POLITE utterance of the 'HELLO' sentence





## C.2.SF-casual: SF's CASUAL utterance of the 'HELLO' sentence





# APPENDIX D

**Acoustic measurements of  $f_0$  and temporal variables of  
utterances spoken by six male speakers**



TABLE D.1. Measurements of f0 variables in polite (P), casual (C) and authoritative (A) versions of two sentences and natural conversations spoken by six male speakers.

(a) 'LUGGAGE' SENTENCE

Speaker	V	Level	Variability		Range		Steepness* <sup>1</sup>
		Mean f0	SD	SD/mean	Range	5%-95%	Mean for
		(Hz)	(Hz)		95%-5%	(Hz)	regression
					(semitone)		coefficients
TN	P	147.1	42.8	0.291	15.9	94 - 236	1.91
	C	155.4	54.5	0.351	18.8	90 - 266	2.61
KS	P	147.0	34.7	0.236	15.5	85 - 208	1.80
	C	161.5	40.3	0.250	14.3	104 - 238	2.21
HA	P	146.1	24.4	0.167	9.4	109 - 188	5.96
	C	135.5	15.1	0.111	5.4	115 - 157	3.99
	A	128.3	14.9	0.116	6.8	106 - 157	2.66
TK	P	123.9	30.1	0.243	11.8	91 - 180	3.16
	C	143.1	26.0	0.182	10.2	102 - 184	4.00
	A	133.8	8.2	0.061	3.1	123 - 147	1.86
SF	P	134.0	37.6	0.280	14.6	90 - 209	3.26
	C	134.3	36.2	0.269	12.8	97 - 203	4.24
KI	P	114.8	23.8	0.208	10.1	91 - 163	3.40
	C	112.5	31.7	0.282	11.4	84 - 162	2.74

V is the speaking style of the version: P is 'polite', C 'casual' and 'A' authoritative.

\*<sup>1</sup>: regression coefficients were calculated with normalised f0 values of each speaker. All f0 values of each speaker were normalised in such a way that the lowest f0 and the highest f0 of the speaker is 0 and 100. The lowest and highest f0 were among f0 values of all the utterances of the two sentences and the natural conversations of the speaker.

## (b) 'HELLO' SENTENCE

<i>Speaker</i>	<i>V</i>	<i>Level</i>	<i>Variability</i>		<i>Range</i>		<i>Steepness</i> <sup>*1</sup>
		<i>Mean f0</i> (Hz)	<i>SD</i> (Hz)	<i>SD/mean</i>	<i>Range</i> 95%-5% (semitone)	<i>5%-95%</i> (Hz)	<i>Mean for</i> <i>regression</i> <i>coefficients</i>
TN	P	163.2	44.1	0.270	15.21	86 - 207	1.49
	C	166.2	50.5	0.304	19.96	78 - 247	1.92
KS	P	170.9	36.3	0.212	11.75	105 - 207	1.57
	C	178.4	43.7	0.245	12.98	112 - 237	1.63
HA	P	156.0	23.3	0.150	9.71	105 - 184	3.16
	C	138.8	28.2	0.203	10.54	99 - 182	3.36
	A	144.0	25.3	0.175	10.51	97 - 178	2.71
TK	P	158.0	20.9	0.132	7.81	114 - 179	2.35
	C	149.1	19.1	0.128	7.02	112 - 168	1.00
	A	150.1	12.9	0.086	5.26	121 - 164	0.98
SF	P	145.1	27.2	0.187	10.99	97 - 183	2.52
	C	147.9	22.1	0.150	9.57	107 - 186	1.79
KI	P	122.1	24.7	0.202	8.77	91 - 151	2.85
	C	128.7	28.5	0.222	10.32	92 - 167	3.21

*V* is the speaking style of the version: P is 'polite', C 'casual' and 'A' authoritative.

<sup>\*1</sup>: regression coefficients were calculated with normalised f0 values of each speaker. All f0 values of each speaker were normalised in such a way that the lowest f0 and the highest f0 of the speaker is 0 and 100. The lowest and highest f0 were among f0 values of all the utterances of the two sentences and the natural conversations of the speaker.



(c) NATURAL CONVERSATIONS

<i>Speaker</i>	<i>Level</i>	<i>Variability</i>		<i>Range</i>	
	<i>Mean f0</i>	<i>SD</i>	<i>SD/mean</i>	<i>Range</i>	<i>5%-95%</i>
	<i>(Hz)</i>	<i>(Hz)</i>		<i>95%-5%</i>	<i>(Hz)</i>
				<i>(semitone)</i>	
TN	122	25.8	0.21	12.19	89 - 180
KS	124	17.4	0.14	7.75	101 - 158
HA	131	18.6	0.14	7.82	105 - 165
TK	103	17.4	0.17	8.37	82 - 133
SF	121	19.0	0.16	7.82	98 - 154
KI	111	19.9	0.18	8.88	88 - 147

TABLE D.2. Measurements of temporal variables in polite (P), casual (C) and authoritative (A) versions of two sentences spoken by six male speakers.

(a) 'LUGGAGE' SENTENCE

<i>Speaker</i>	<i>V</i>	<i>Speech rate</i> <sup>*1</sup> ( <i>mora/sec</i> )	<i>Total utterance</i> <sup>*2</sup> ( <i>ms</i> )	<i>Phrase 1</i> ( <i>ms</i> )	<i>Pause</i> ( <i>ms</i> )	<i>Phrase 2</i> ( <i>ms</i> )
TN	P	10.4	1280	470	70	740
	C	11.6	1250	450	70	730
KS	P	10.8	1120	340	50	730
	C	11.9	1140	310	20	810
HA	P	11.6	1030	360	0	670
	C	14.3	970	270	0	700
	A	15.2	920	290	0	630
TK	P	11.8	1050	360	10	680
	C	14.1	960	270	0	690
	A	13.3	980	290	0	690
SF	P	11.8	1140	430	110	600
	C	12.3	970	330	0	640
KI	P	11.8	1060	300	10	750
	C	13.2	950	280	0	670

*V* is the speaking style of the version: P is 'polite', C 'casual' and 'A' authoritative.

<sup>\*1</sup>: the duration of a pause between Phrase 1 and Phrase 2, and the final morae in Phrase 1 and Phrase 2 were excluded from calculation of speech rate.

<sup>\*2</sup>: the total utterance consists of Phrase 1, a pause and phrase 2.



(b) 'HELLO' SENTENCE

<i>Speaker</i>	<i>V</i>	<i>Speech rate*<sup>1</sup></i> <i>(mora/sec)</i>	<i>Total utterance*<sup>2</sup></i> <i>(ms)</i>	<i>Phrase 1</i> <i>(ms)</i>	<i>Pause</i> <i>(ms)</i>	<i>Phrase 2</i> <i>(ms)</i>
TN	P	11.5* <sup>3</sup>	2160	510	520	1130
	C	13.2	2180	460	700	1020
KS	P	11.9	2000	510	400	1090
	C	13.2	1710	430	110	1170
HA	P	13.5	1880	430	500	950
	C	13.9	1930	410	490	1030
	A	14.1	2310	440	860	1010
TK	P	12.7	1530	380	160	990
	C	13.0	1590	350	150	1120
	A	13.0	1590	390	150	1050
SF	P	13.5	1500	410	150	940
	C	14.1	1320	340	0	980
KI	P	12.8	1490	380	0	1110
	C	13.0	1440	350	0	1090

*V* is the speaking style of the version: P is 'polite', C 'casual' and 'A' authoritative.

\*<sup>1</sup>: the duration of a pause between Phrase 1 and Phrase 2, and the final morae in Phrase 1 and Phrase 2 were excluded from calculation of speech rate.

\*<sup>2</sup>: the total utterance consists of Phrase 1, a pause and phrase 2.

\*<sup>3</sup>: since the most natural utterance by TN for politeness had a slight tongue twist in the middle, the second best utterance was used for calculation of speech rate.

TABLE D.3. Measurements of the duration and steepness of the final morae in polite (P), casual (C) and authoritative (A) versions of two sentences spoken by six male speakers.

(a) 'LUGGAGE' SENTENCE

<i>Speaker</i>	<i>V</i>	<i>Final mora in Phrase 1</i>		<i>Final vowel in Phrase 2</i>
		<i>Duration</i>	<i>Steepness</i>	<i>Duration</i>
		<i>(ms)</i>	<i>(semitone/s)</i>	<i>(ms)</i>
TN	P	160	-53	140
	C	50	-111	160
KS	P	70	-83	110
	C	70	-104	240
HA	P	60	-119	60
	C	80	-125	150
	A	100	-36	130
TK	P	90	-68	70
	C	70	-90	130
	A	70	-45	110
SF	P	160	-61	60
	C	70	(-160) + (-42)	70
KI	P	60	-12	110
	C	50	-93	80

*V* is the speaking style of the version: P is 'polite', C 'casual' and 'A' authoritative.



(b) 'HELLO' SENTENCE

<i>Speaker</i>	<i>V</i>	<i>Final mora in Phrase 1</i>		<i>Final vowel in Phrase 2</i>
		<i>Duration</i>	<i>Steepness</i>	<i>Duration</i>
		<i>'sh' + 'i'</i> (ms)	(semitone/s)	(ms)
TN	P	110 + 130	-73	120
	C	90 + 150	-63	120
KS	P	100 + 120	(-75) + (-11)	160
	C	80 + 90	-59	260
HA	P	120 + 90	-25	70
	C	110 + 100	-14	150
	A	80 + 120	4	180
TK	P	130 + 0	devoiced	50
	C	130 + 0	devoiced	190
	A	160 + 0	devoiced	130
SF	P	110 + 80	-69	80
	C	70 + 40	-58	130
KI	P	110 + 50	-29	120
	C	90 + 30	-74	130

*V* is the speaking style of the version: P is 'polite', C 'casual' and 'A' authoritative.

# APPENDIX E

## Instructions given to subjects in Experiment 1

E.1. Original instructions (in Japanese)

E.2. English translation of E.1

E.3. Written text for measurements  
of speech rate of subjects



## E.1. Original instructions

TAPE-ID: \_\_\_\_\_

LISTENER-JUDGE:

Sex: Male Female

Age: \_\_\_\_\_

Hometown: \_\_\_\_\_

Dialect: \_\_\_\_\_

(方言)

## INSTRUCTIONS:

これからお聞きになるテープには、ただ1つの文章「荷物はこれだけですが」が様々な声の調子で話されたものが入っています。これはもともと税関の人がいろいろなタイプの乗客に質問しているという想定で録音されました。

それぞれの発話は、ビーという音で始まり、短いサイレンスが続きますので、この間に、ご自分の判断基準に従って 丁寧度 (Politeness) 相手に対してスピーカーがどのくらい丁寧に話しているか) をつけてください。

考えはじめるとわからなくなりますので、なるべく直観的につけるようにしてください。

## PRACTICE:

	VERY IMPOLITE 非常に丁寧でない	VERY POLITE 非常に丁寧
[P1]	-----+-----	
[P2]	-----+-----	
[P3]	-----+-----	
[P4]	-----+-----	
[P5]	-----+-----	
[P6]	-----+-----	

## Instructions (2)

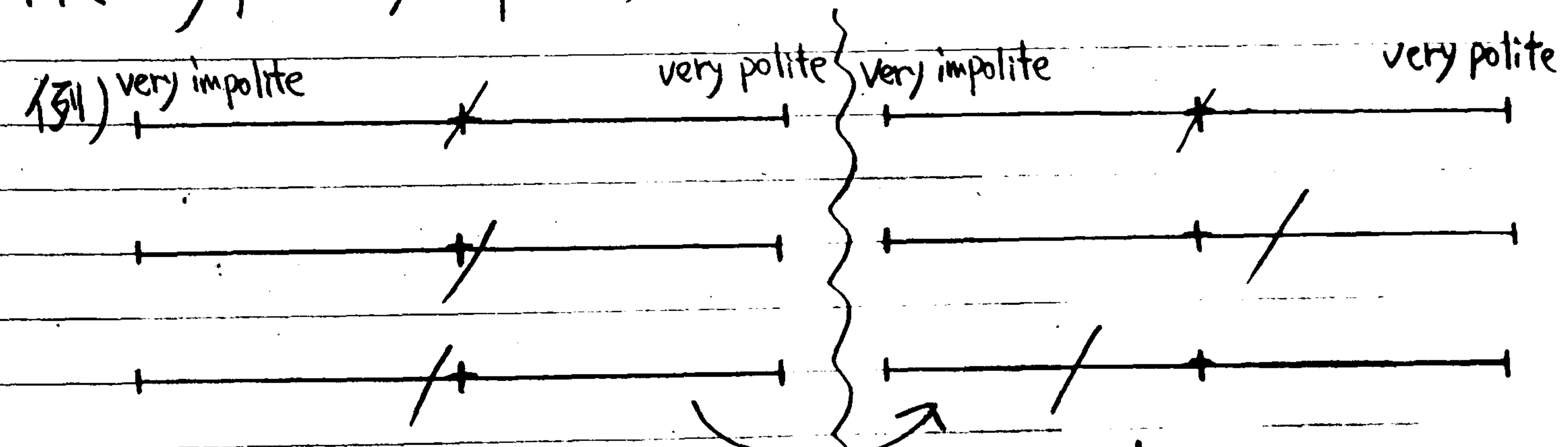
これから聞いていただく発話は 2人の話し手が 税関の若い役人の立場になって 丁唯いに (立派な地位のある紳士に向かって、ちょっとかしまって) 言っているものと、丁唯いでなく (自分と同じくらいの年がこの学生に向かって 気楽に) 言っているものです。

すべての発話は、どれも「非常に」丁唯い/丁唯いでないというものではありません。従ってどちらかという微妙な違いを聞いていただくこととなります。

スケールの中間の「+」が、丁唯いさの軸の中間で、発話がどのくらい④にあるいは⑤に感じられるか、特に同じ④/⑤に感じられる中でも どの程度そう感じられるかを、大胆につけて下さい。

これから まず 4つの発話を聞いていただきますが、最初の2つが、この話し手たちが、丁唯いに話したものと、あとの2つが、気楽に話したものです。すべてが、ご自分の内部基準では「少し」丁唯い/丁唯いでないの範囲に入ってしまうかもしれませんが、なるべくこの answer sheet のスケールの

枠 (very polite / impolite) 一杯に拡大してつけるようにお願いします。



もう少し大きくふって下さい!  
(もし違いがあると感じられたなら)



E.2. English translation of E.1

TAPE-ID: \_\_\_\_\_

LISTENER-JUDGE:

Sex: Male Female

Age: \_\_\_\_\_

Hometown: \_\_\_\_\_

Dialect: \_\_\_\_\_

INSTRUCTIONS:

You will hear only one Japanese sentence 'Nimotsu-wa koredake desuka' spoken in various ways. The utterances are recordings of hypothetical situations in which a customs officer is talking to various types of passengers.

Each utterance is preceded by a beep tone and followed by a short period of silence during which you are asked to rate the utterance on a scale of *politeness* (how politely the speaker is speaking to the addressee) according to your own criteria.

Do not think carefully, just rate them intuitively.

PRACTICE:

	VERY IMPOLITE	VERY POLITE
[P1]	-----+-----	
[P2]	-----+-----	
[P3]	-----+-----	
[P4]	-----+-----	
[P5]	-----+-----	
[P6]	-----+-----	

## INSTRUCTIONS (2)

The utterances you are going to hear were spoken by two speakers who were asked to imagine themselves as a young customs officer talking to a respectable gentleman politely, and a young student casually. None of the utterances are 'VERY polite/impolite'. Therefore, the difference between the utterances may be rather subtle. The sign '+' on the politeness scale indicates the neutral point. You are asked to rate each utterance on this politeness scale, by judging how politely or how impolitely the utterance sounded to you. Rate them on the scale in a decisive way.

First, you are going to hear four utterances. The first two are utterances which were meant to be polite by the two speakers, and the latter two were meant to be casual. Although all the four utterances might fall in the range between slightly impolite and slightly polite by your standard, try to magnify this range to the entire scale on this answer sheet (Very impolite - very polite).

e.g.,

VERY impolite    VERY polite            VERY impolite    VERY polite

|-----/-----|    ->    |-----/-----|

|-----+/-----|    ->    |-----+---/-----|

|-----/+-----|    ->    |-----/---+-----|



**E.3. Written text for measurements of speech rate of subjects**

こんにちは、はじめまして。  
あの、すみませんが、何も書く物を  
持ってこなかったのて、そのボールペン  
を貸してもらえますか。

(English translation)

"Hello, how do you do.  
erm, I'm afraid I haven't brought anything to write with,  
could I borrow the ball-point pen over there, please?"

## **APPENDIX F**

**Waveforms and f0 contours of the 'angry' and 'kind' source utterances and 3D plot of the final mora of these utterances used in Experiment 2, originally spoken by a trained male speaker**

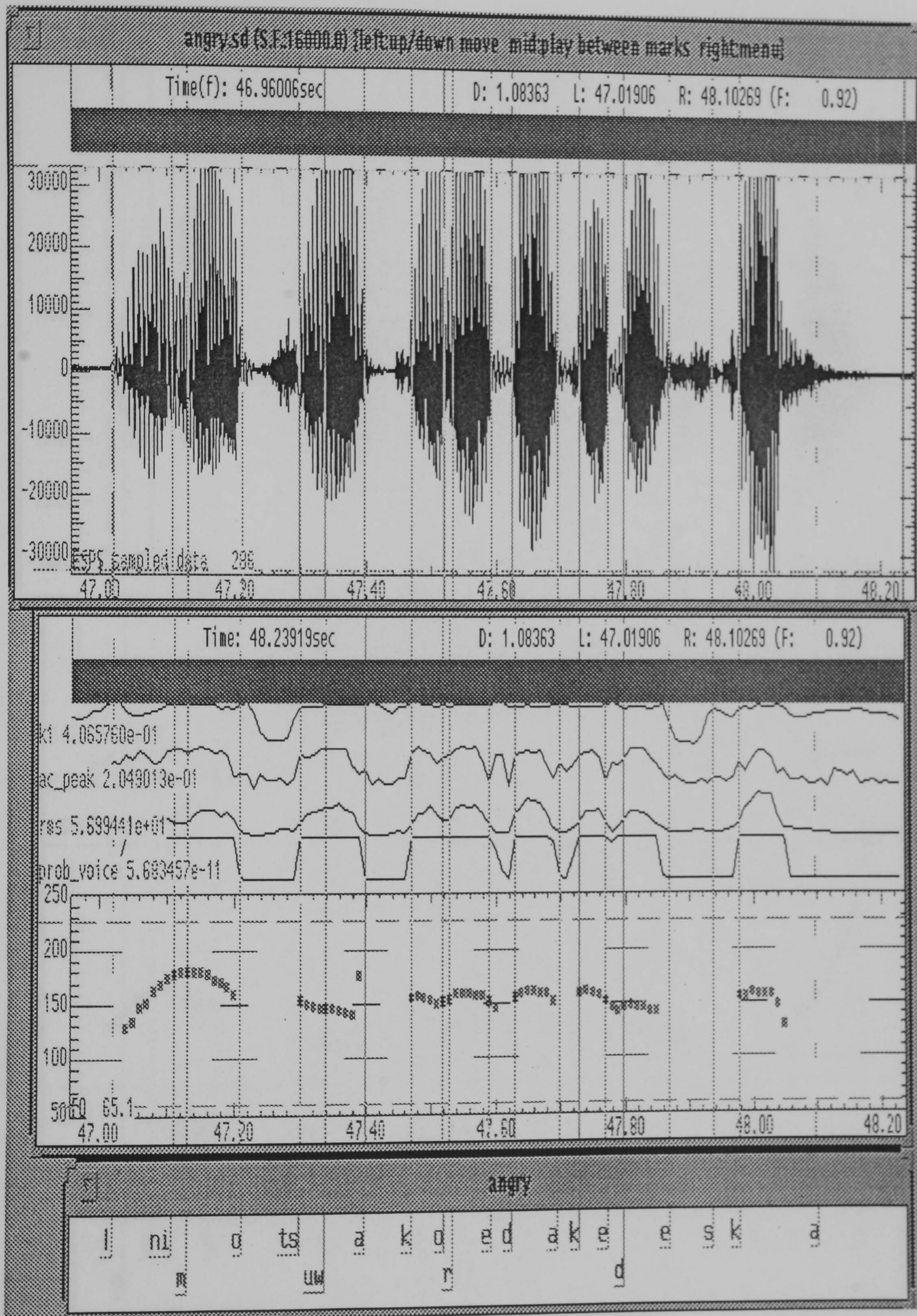
F.1. Waveforms and f0 contours  
of the 'angry' and 'kind' utterances

F.2. 3D plot of the final mora 'ka'



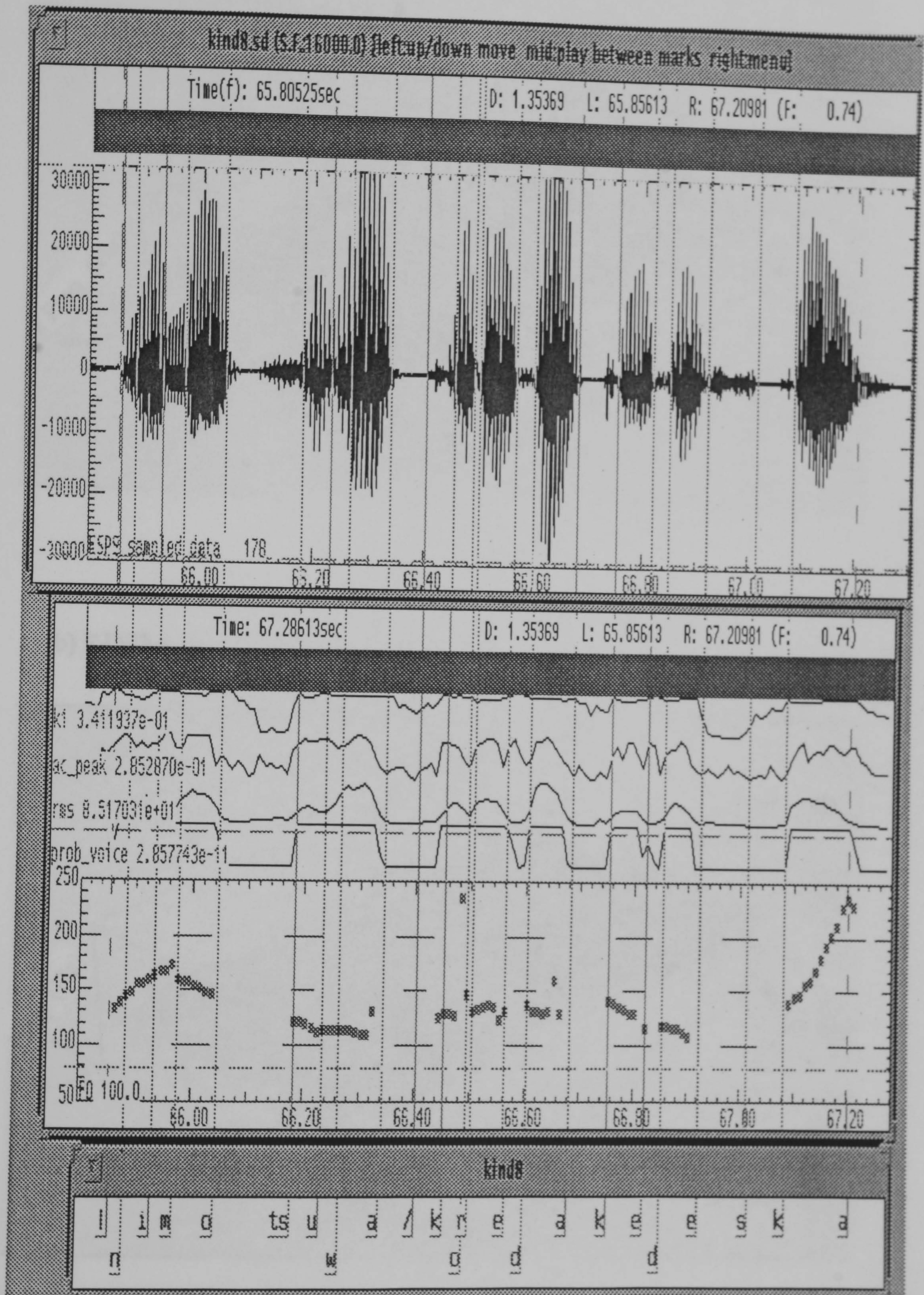
# F.1. Waveforms and f0 contours of the 'angry' and 'kind' utterance

(a) ANGRY style





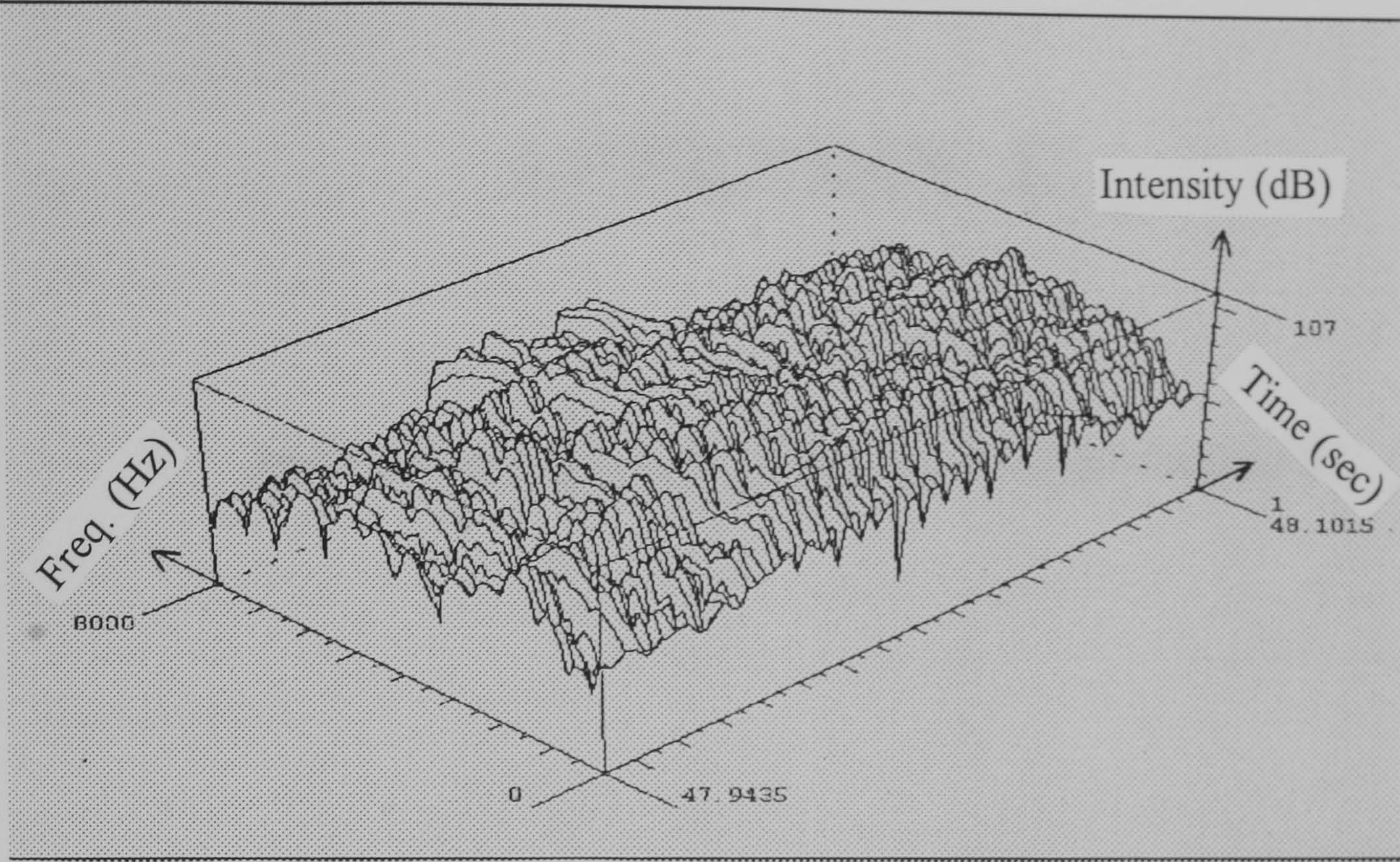
(b) KIND style



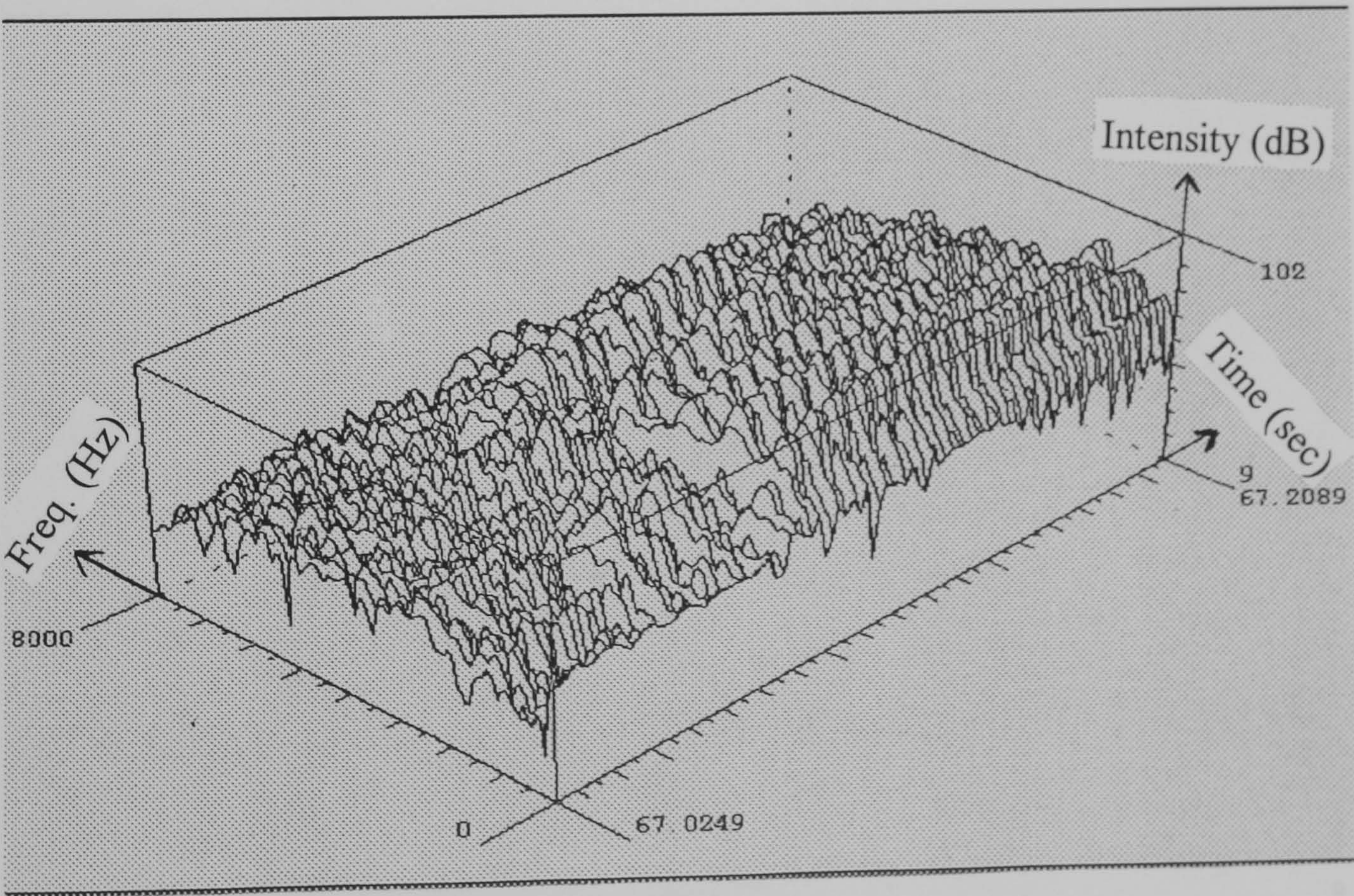


F.2. 3D plot of the final mora 'ka'

(a) ANGRY style



(b) KIND style





# **APPENDIX G**

## **Instructions given to subjects in Experiment 2**



SUBJECT: \_\_\_\_\_  
 SEX: F M  
 AGE: \_\_\_\_\_  
 HOMETOWN or DIALECT: \_\_\_\_\_

### Operation

- > 1. When you are ready, press ENTER  
 [2] An utterance through the loudspeaker  
 3. Rate the utterance on the four scales  
 \_\_\_\_ [4] When you finish marking,  
           the EXPERIMENTER presses ENTER

### Instructions for rating

- + Rate each utterance on ALL four scales
- + The most APPROPRIATE scale FIRST
- + For each utterance,
  - (1) CODE: on what scale  
       Select a code from  
       P(olite), A(ngry), K(ind), and N(atural)
  - (2) SCALE: the degree  
       Mark the scale with 'v'

### Example:

Number	CODE	SCALE
		-VERY                  NEUTRAL                  +VERY
10	[K]	-----0---v-----
	[P]	-----0---v-----
	[N]	-----0-v-----
	[A]	-----v-----
11	[N]	--v-----0-----
	[P]	-----v-----0-----
	[A]	-----v-----
	[K]	-----0-v-----

/\*\* Practice session \*\*/

NUMBER CODE		SCALE		
		-VERY	NEUTRAL	+VERY
[ ]	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		
[ ]	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		
[ ]	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		
[ ]	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		
	[ ]	-----0-----		



# APPENDIX H

## Instructions given to subjects in Experiment 3

(used for both politeness/naturalness sessions)

You are going to hear only one sentence in this session.

*Nimotsu-wa koredake desuka*

The speakers were given some scenarios, which described the situations, and asked to say this sentence appropriately in a given situation.

Setting: at a customs counter at an airport

Speaker: playing a role of a young customs officer

Addressee:

- (A) a respectable gentleman
- (B) a young casually dressed student
- (C) a drunk/trouble maker

The content of this tape is:

#### DAY 1

SECTION 0: practice session (4 pairs)

SECTION 1: 38 pairs  
beep beep

SECTION 2: 60 pairs  
beep beep

SECTION 3: 60 pairs  
beep beep

SECTION 4: 60 pairs  
beep beep beep

#### DAY 2

SECTION 1: 60 pairs  
beep beep

SECTION 2: 60 pairs  
beep beep

SECTION 3: 60 pairs  
beep beep

SECTION 4: 60 pairs  
beep beep beep

Each pair consists of

- beep
  - Utterance 1
  - Utterance 2
  - 2-second silence
- during which you are asked to make a judgement



(for politeness sessions)

ANSWER SHEET

TAPE-ID: \_\_\_\_\_

SEX: Male Female

AGE: \_\_\_\_\_

HOMETOWN/DIALECT: \_\_\_\_\_

Instruction:

Encircle the utterance (1st or 2nd) which sounded more POLITE to you.

SECTION 0		
-----		
Practice session		
-----		
[P1]	1st	2nd
[P2]	1st	2nd
[P3]	1st	2nd
[P4]	1st	2nd
-----		

「丁寧さは状況によって様々です  
が、ここでは若い親類の役人が  
きちんとした身なりの紳士に  
(敬意をもって)話しているという  
状況における「丁寧さ」を  
考えてください。

(English translation)

"Politeness varies from situation to situation. The situation here is that a young customs officer is speaking to a respectable gentleman (with respect)."

(for naturalness sessions)

ANSWER SHEET

TAPE-ID: \_\_\_\_\_

SEX: Male Female

AGE: \_\_\_\_\_

HOMETOWN/DIALECT: \_\_\_\_\_

Instruction:

Encircle the utterance (1st or 2nd) which sounded more NATURAL to you.

SECTION 0

Practice session

[P1]	1st	2nd
[P2]	1st	2nd
[P3]	1st	2nd
[P4]	1st	2nd

「自然さ」も音質、スピード、言い方  
あるいは状況における適切さなど  
様々なレベルがあり、ここでは  
総合的判断として 2つの発話の  
どちらが「自然」に聞こえるかを判断  
してください。

(English translation)

"There are various levels of naturalness such as speech quality, tempo, the way of speaking and the appropriateness in a given situation. Judge NATURALNESS as a global judgement."



# APPENDIX I

## **Politeness and naturalness scores in Experiment 3**

TABLE I.1. Politeness and naturalness scores of five trial blocks (T1 ~ T5) and the mean value across each subject's scores of the trial blocks for the polite utterances spoken by two speakers, in Experiment 3. Speech rate level is a level of compression/expansion rate in segmental duration of the source utterance: S20 is 20% expansion, S10, 10% expansion, UM, unmodified, F10, 10% compression and F20, 20% compression.

(a-1) SUBJECT: M1    POLITENESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.00	0.00	0.11	0.06	0.06	0.04
	S10	0.22	0.28	0.17	0.33	0.22	0.24
	UM	0.33	0.33	0.33	0.33	0.28	0.32
	F10	0.28	0.39	0.39	0.22	0.39	0.33
	F20	0.28	0.22	0.17	0.22	0.22	0.22
TK	S20	0.72	0.67	0.61	0.56	0.78	0.67
	S10	0.72	0.78	0.78	0.72	0.83	0.77
	UM	0.94	1.00	0.94	1.00	0.89	0.96
	F10	0.89	0.72	0.89	0.83	0.72	0.81
	F20	0.61	0.61	0.61	0.72	0.61	0.63

(a-2) SUBJECT: M1    NATURALNESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.33	0.22	0.22	0.17	0.17	0.22
	S10	0.56	0.44	0.50	0.44	0.56	0.50
	UM	0.72	0.72	0.67	0.72	0.78	0.72
	F10	0.56	0.56	0.56	0.61	0.72	0.60
	F20	0.11	0.22	0.22	0.17	0.28	0.20
TK	S20	0.44	0.67	0.33	0.39	0.39	0.44
	S10	0.56	0.72	0.83	0.78	0.78	0.73
	UM	0.72	0.83	0.94	0.89	0.61	0.80
	F10	0.72	0.50	0.44	0.72	0.56	0.59
	F20	0.28	0.11	0.28	0.11	0.17	0.19



(b-1) SUBJECT: M2    POLITENESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.11	0.22	0.28	0.39	0.33	0.27
	S10	0.50	0.44	0.50	0.44	0.39	0.46
	UM	0.61	0.61	0.39	0.50	0.56	0.53
	F10	0.56	0.17	0.33	0.39	0.39	0.37
	F20	0.17	0.11	0.06	0.17	0.11	0.12
TK	S20	0.67	0.72	0.78	0.61	0.72	0.70
	S10	0.67	0.78	0.89	0.89	0.83	0.81
	UM	0.94	0.94	0.89	0.72	0.83	0.87
	F10	0.50	0.67	0.50	0.67	0.61	0.59
	F20	0.28	0.33	0.39	0.22	0.22	0.29

(b-2) SUBJECT: M2    NATURALNESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.11	0.00	0.22	0.22	0.11	0.13
	S10	0.50	0.39	0.50	0.44	0.61	0.49
	UM	0.67	0.83	0.72	0.72	0.50	0.69
	F10	0.61	0.72	0.67	0.61	0.61	0.64
	F20	0.44	0.61	0.50	0.33	0.17	0.41
TK	S20	0.22	0.28	0.06	0.22	0.33	0.22
	S10	0.78	0.44	0.67	0.67	0.72	0.66
	UM	0.67	0.72	0.67	0.78	0.78	0.72
	F10	0.78	0.67	0.72	0.72	0.78	0.73
	F20	0.22	0.33	0.28	0.28	0.39	0.30

(c-1) SUBJECT: F1    POLITENESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.11	0.11	0.22	0.00	0.06	0.10
	S10	0.17	0.17	0.17	0.22	0.17	0.18
	UM	0.44	0.28	0.22	0.39	0.39	0.34
	F10	0.28	0.33	0.22	0.28	0.22	0.27
	F20	0.33	0.28	0.28	0.28	0.33	0.30
TK	S20	0.44	0.78	0.72	0.61	0.72	0.66
	S10	0.83	0.72	0.72	0.78	0.89	0.79
	UM	0.83	0.78	0.83	0.67	0.72	0.77
	F10	0.72	0.78	0.94	0.83	0.78	0.81
	F20	0.83	0.78	0.67	0.94	0.72	0.79

(c-2) SUBJECT: F1    NATURALNESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.44	0.33	0.50	0.56	0.39	0.44
	S10	0.61	0.67	0.72	1.00	0.72	0.74
	UM	0.83	0.83	0.83	0.83	0.89	0.84
	F10	0.83	0.83	0.78	0.72	0.78	0.79
	F20	0.50	0.61	0.39	0.67	0.61	0.56
TK	S20	0.17	0.11	0.22	0.22	0.28	0.20
	S10	0.33	0.22	0.33	0.28	0.28	0.29
	UM	0.50	0.56	0.44	0.33	0.44	0.46
	F10	0.50	0.44	0.44	0.22	0.28	0.38
	F20	0.28	0.39	0.33	0.17	0.33	0.30



(d-1) SUBJECT: F2    POLITENESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.06	0.00	0.00	0.06	0.06	0.03
	S10	0.17	0.11	0.22	0.11	0.17	0.16
	UM	0.56	0.44	0.61	0.67	0.50	0.56
	F10	0.89	0.78	0.89	0.78	0.89	0.84
	F20	0.67	0.83	0.94	1.00	0.94	0.88
TK	S20	0.17	0.22	0.22	0.28	0.28	0.23
	S10	0.44	0.50	0.33	0.28	0.22	0.36
	UM	0.44	0.50	0.39	0.50	0.56	0.48
	F10	0.89	0.67	0.61	0.61	0.56	0.67
	F20	0.72	0.94	0.78	0.72	0.83	0.80

(d-2) SUBJECT: F2    NATURALNESS SCORES

<i>Speaker</i>	<i>Speech rate level</i>	<i>Scores</i>					
		<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>Mean</i>
KS	S20	0.06	0.00	0.22	0.00	0.06	0.07
	S10	0.33	0.33	0.33	0.17	0.11	0.26
	UM	0.72	0.33	0.39	0.50	0.61	0.51
	F10	0.72	0.67	0.67	0.72	0.67	0.69
	F20	0.72	0.78	0.72	0.78	0.72	0.74
TK	S20	0.06	0.22	0.00	0.17	0.22	0.13
	S10	0.28	0.33	0.44	0.33	0.50	0.38
	UM	0.61	0.56	0.61	0.61	0.50	0.58
	F10	0.83	0.83	0.78	0.78	0.83	0.81
	F20	0.67	0.94	0.83	0.94	0.78	0.83

# APPENDIX J

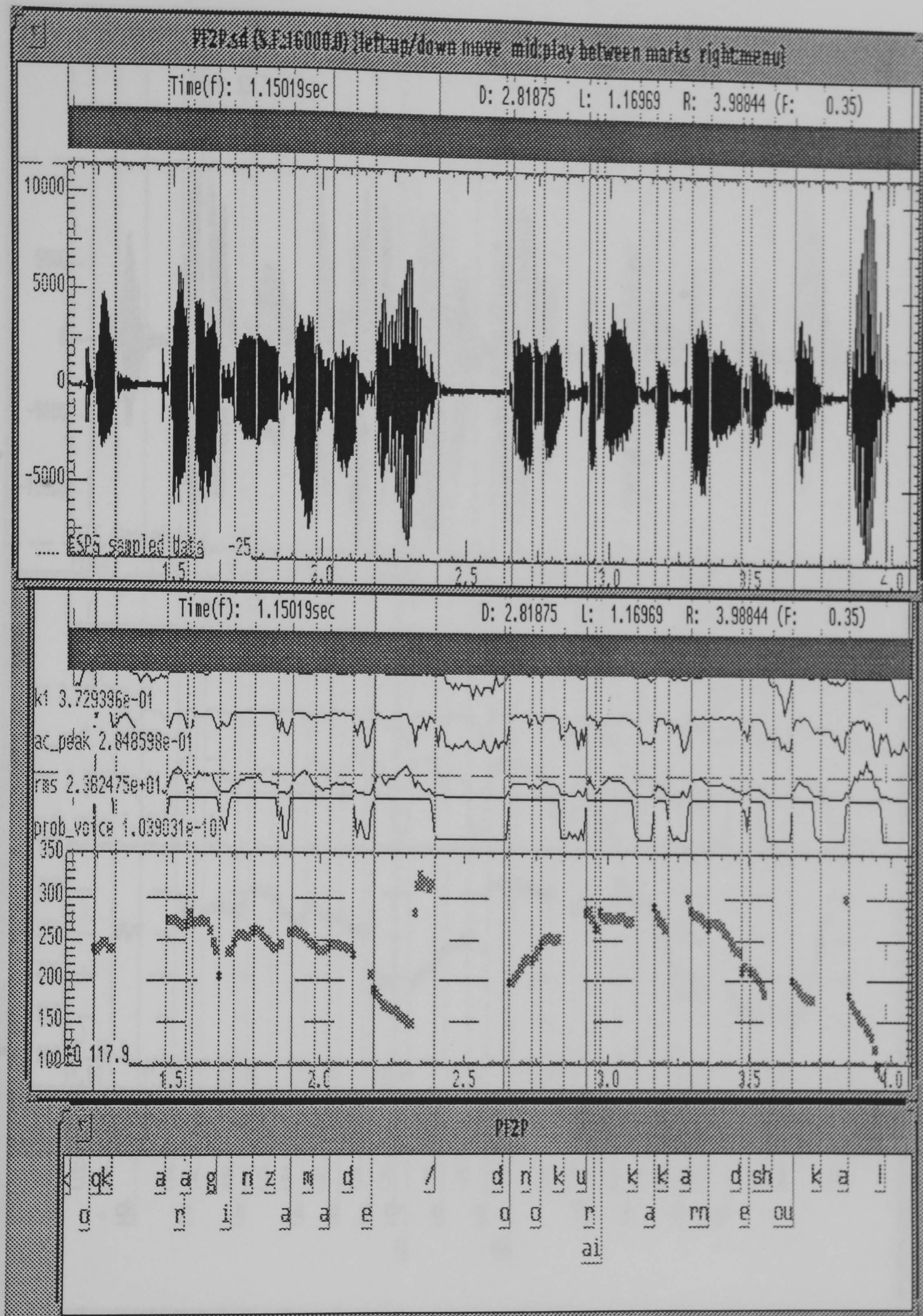
## **Waveforms and f0 contours of polite and non-polite utterances by a human speaker and the SYNCON synthesiser**

J.1. Utterances of a trained female speaker (PF2)  
(PF2P and PF2I)

J.2. Utterances produced by SYNCON  
(D(P)F0(P) and D(I)F0(I))

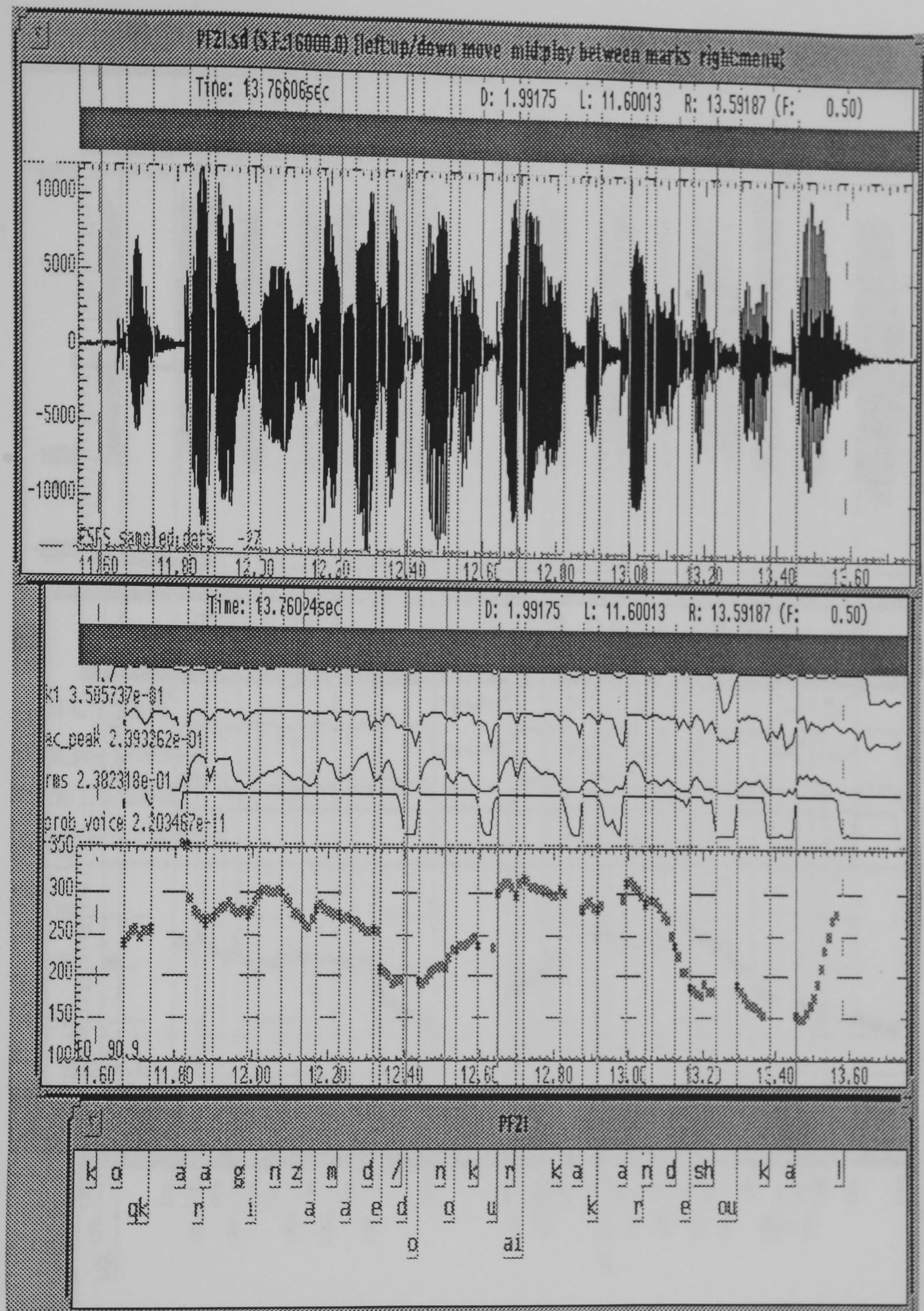


# J.1.PF2P: Original polite utterance by a trained female speaker



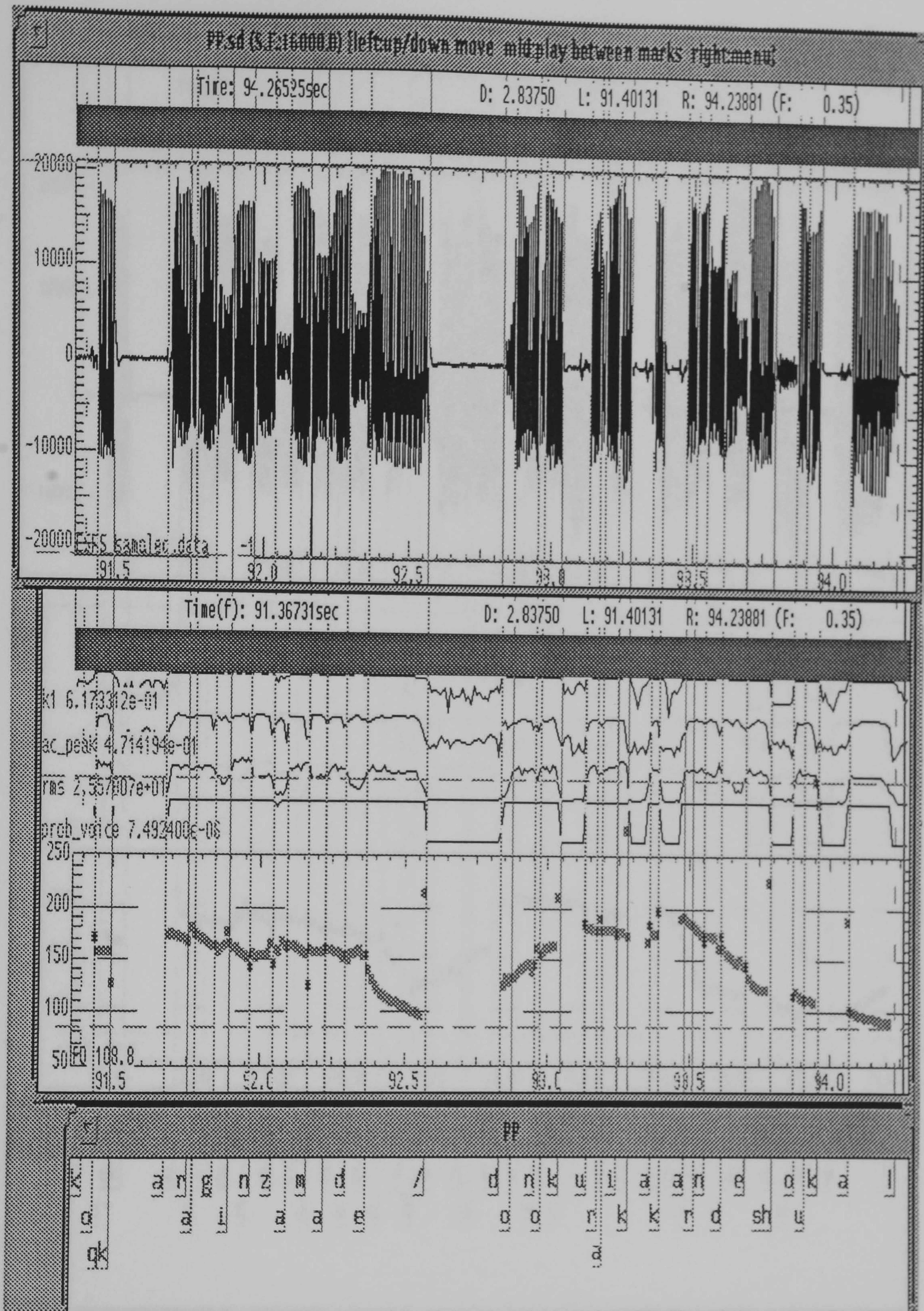


# **J.1.PF2I: Original non-polite utterance by a trained female speaker**





# J.2.D(P)F0(P): Synthetic polite utterance





# J.2.D(I)F0(I): Synthetic non-polite utterance

